



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/82, 15/29, A01H 5/00</b>	<b>A1</b>	(11) International Publication Number: <b>WO 98/45458</b> (43) International Publication Date: 15 October 1998 (15.10.98)
<p>(21) International Application Number: PCT/US98/06673</p> <p>(22) International Filing Date: 6 April 1998 (06.04.98)</p> <p>(30) Priority Data: 60/042,827 8 April 1997 (08.04.97) US</p> <p>(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).</p> <p>(72) Inventor; and (75) Inventor/Applicant (for US only): GUTTERIDGE, Steven [US/US]; 4 Austin Road, Wilmington, DE 19810 (US).</p> <p>(74) Agent: CHRISTENBURY, Lynne, M.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).</p>	<p>(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GW, HU, ID, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> With international search report.</p>	
<p>(54) Title: AN ENGINEERED SEED PROTEIN HAVING A HIGHER PERCENTAGE OF ESSENTIAL AMINO ACIDS</p> <pre>                                 NdeI GGGGAACCTTCTATGTCAGGAGGAGTGTCCGAGCAGATCGACAGACAGCAGATGCTCAG 1  -----+-----+-----+-----+-----+-----+-----+-----+ CCCTTGGAGGTATACGCTCTCCTCAGCGCTCGTCTACGCTCTCTGTCTTACGAGTC      60                                 M Q E E C R E Q M Q R Q Q M L S CCACTGCCGATGTACATGAGACAGCAGATGGAGGAGAGCCCGTACCGAGCATGCCAG 61  -----+-----+-----+-----+-----+-----+-----+-----+ GGTGACGGCTACATGTACTCTGTCTCTA C T C T C T C G G C A T G G T C T G G T A C G G G T C H C R M Y M R Q Q M E E S P Y Q T M P R GCGGGAATGGAGCCGACATGAGCGAGTGCTCGCAGAGCTGGAGGGGATGGACGAGAG 121 -----+-----+-----+-----+-----+-----+-----+-----+ CGCCCTTACCTCGGCGTGTACTCGCTCAGGAGCTGTGTGAGCTCCCTACCTGCTCTC R G M E P H M S E C C E Q L E G M D E S                                 PetI CTGCAGATCGGAAGGCTTAAGGATGATGATGATGAGGATGCAACAGGAGGAGATGCAACC 181 -----+-----+-----+-----+-----+-----+-----+-----+ GACGCTACGCTTCGGAATCTCTACTACTACTCTCTACGTTGTCTCTCTCTACGTTGG C R C E G L R M M M M R M Q Q E E M Q P CCGAGGGGAGCAGATGCGAAGGATGATGAGGCTGGCCGAGAATATCCCTCCCGCTGCAA 241 -----+-----+-----+-----+-----+-----+-----+-----+ GGCTCCCTCGTCTACGCTTCTCTACTCTCGACCGGCTCTTATAGGGAAGGGGACGCTT R G E Q M R R M M R L A E N I P S R C N </pre>		
<p>(57) Abstract</p> <p>This invention pertains to development of seeds and seed storage proteins that are enhanced in the quantity of amino acids that are essential to humans and animals. More specifically, this invention concerns the genetic engineering of the Brazil Nut 2S albumin seed storage protein so that it contains a higher percentage of essential amino acid residues. Expression of a gene encoding this engineered seed storage protein in transgenic plants results in increased accumulation of essential amino acids in the seeds of these plants.</p>		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## TITLE

### AN ENGINEERED SEED PROTEIN HAVING A HIGHER PERCENTAGE OF ESSENTIAL AMINO ACIDS

## FIELD OF THE INVENTION

5 This invention pertains to the development of seeds and seed storage proteins that are enhanced in the quantity of amino acids that are essential to humans and animals, and more particularly, the enhancement of the quantity of essential amino acids in the Brazil Nut 2S albumin seed storage protein.

## BACKGROUND OF THE INVENTION

10 Many vertebrates, including man, lack the ability to manufacture a number of amino acids, and therefore, require these preformed in the diet. These are called essential amino acids. The two major sources of dietary protein in the US, corn and soybeans, are deficient in some nutritionally indispensable (essential) amino acids. The amino acids essential for humans and most animals that must be acquired from dietary  
15 sources include the sulfur-containing residues methionine (Met) and cysteine (Cys), along with the basic amino acid lysine (Lys) and aromatic tryptophan (Trp). Soybean meal is a good source of Lys and Trp but poor in sulfur-containing residues and thus must be supplemented with sulfur-rich corn meal to provide a suitably balanced diet. A protein that has a substantial proportion of both the sulfur-amino acids and Lys in  
20 content that can be expressed to high levels in seeds of crop plants would have two advantages. First, the need to supplement meals with individual amino acids, or blend different meals would be obviated. Second, other meals that are left after extraction of other commodities and presently discarded for lack of nutrition, might become  
25 alternative sources of balanced dietary protein. With the molecular genetic tools now available, alteration of the amino acid composition of seed storage proteins to enhance their nutritional quality is possible. Such altered seed storage proteins can in turn enhance grain amino acid composition, thus adding value for the farmer.

Efforts to improve the amino acid content of crops through plant breeding have resulted in only limited success and then only in the laboratory. A mutant corn line with  
30 elevated whole kernel methionine concentration was isolated from cell culture after selecting for growth in inhibitory concentrations of lysine and threonine (Phillips et al., (1985) *Cereal Chem.* 62:213-218). Similarly, soybean cell lines with increased intracellular concentrations of methionine selected with ethionine have been reported (Madison and Thompson (1988) *Plant Cell Reports* 7, 472-476), but no plants were  
35 regenerated from these lines.

The amino acid content of seeds is determined primarily by the storage proteins synthesized during seed development that serve as a major nutrient reserve following

germination. The quantity of this reserve varies from about 10% dry weight in cereals to 40% in legumes. In some seeds, storage proteins can account for 50% or more of the total protein. Although this abundance has meant that these proteins were some of the first to be isolated, it is only recently that their amino acid sequences have been

determined. A number of sulfur-rich plant seed storage proteins have been identified and their genes isolated. A gene from corn coding for a 15kDa zein protein containing 11% methionine and 5% cysteine (Pedersen et al., (1986) *J. Biol. Chem.* 261, 6279-6284) and one coding for a 10kDa zein containing 23% methionine and 3% cysteine have been isolated (Kiriwara et al., (1988) *Mol. Gen. Genet.* 21, 477-484; Kiriwara et al., *Gene* 71, 359-370), as well as another zein containing 37% methionine and 3% cysteine (Chui and Falco, *Plant Physiol.* 107:291, 1995). Two seed albumin genes from pea containing 8% and 16% cysteine have been reported (Higgins et al., (1986) *J. Biol. Chem.* 261, 11124-11130). The gene from Brazil Nut for a seed 2S albumin containing 18% methionine and 8% cysteine has been isolated (Altenbach et al., (1987) *Plant Mol. Biol.* 8, 239-250). Finally, a gene from rice codes for a 10kDa seed prolamin that has 19% methionine and 10% cysteine (Masumura et al., (1989) *Plant Mol. Biol.* 12, 123-130). Combining the genetic signals controlling expression and targeting in the seed with an engineered storage protein that has enhanced amino acid composition, would provide an attractive means of altering seed protein quality.

The use of natural variants of seed proteins rich in essential amino acids is a promising approach, but applicable only when natural variants rich in the desired amino acid can be found. Few natural proteins with high lysine content have been identified. Proteins with combinations of essential amino acids are even less common, particularly ones with sufficiently high percentages of, for example, methionine and lysine so that the expression levels required to raise the level of those amino acids in seeds still expressing endogenous proteins are not beyond the limits of gene expression technology. Modern protein engineering technology offers a route to create such proteins. One solution is to design proteins completely de novo, such as taught by Falco et al. (World Patent Publication No. WO93/03160). This strategy is risky in that the fate of such a protein in the cell is difficult to predict.

An alternative approach is to re-engineer a pre-existing storage protein already replete with at least one of the essential amino acids. The issues surrounding the expression of modified seed proteins have been reviewed (Krebbes et al. In: *Transgenic Plants*, A. Hiatt, ed. Dekker Inc., New York, 1993, pp 37-60). Briefly, such modifications must not disrupt the complex folding, processing, or intracellular transport processes that these proteins undergo; as exemplified by Hoffman et al. (1989)

*Plant Mol. Biol.* 11, 717-729) when due care is not taken the resulting modified protein risks being degraded.

The Brazil Nut 2S albumin represents a family of related proteins found in a variety of species (Youle and Huang (1981) *Amer. J. Bot.* 68:44-48). They are small proteins found *in vivo* as two subunits linked by disulfide bridges. The two subunits are derived from a single precursor peptide which is extensively processed (Crouch et al. (1983) *J. Mol. App. Genet.* 2, 273-283; Krebbers et al. (1988) *Plant Physiol.* 87, 859-866). Sequence analysis of 2S albumins from different species shows that while the sequences from different species are not always highly conserved, the number of cysteine residues and their arrangement in the sequence is, suggesting that the structure of 2S albumins is similar between species.

There are many reports detailing the expression of the 2S albumin from Brazil Nut in transgenic plants. For example, it has been expressed in the seeds of transformed tobacco under the control of the phaseolin promoter. The 17 kDa precursor form of the protein was correctly processed to the mature dimeric state composed of a 9 kDa and 3 kDa subunit. The accumulation in the seeds of the tobacco resulted in an increase of about 30% in methionine in the seeds (Altenbach et al. (1989) *Plant Mol. Biol.* 13, 513-522). With varying degrees of success the same protein has been expressed in *Brassica*, *Arabidopsis*, and *Vicia narbonensis* and soybean, both grain legumes (Altenbach, (1992) *Plant Mol. Biol.* 18, 235-245; Guerche et al., (1990) *Mol. Gen. Genetics* 221, 306-314; De Clercq et al. (1990) *Plant Physiol.* 94, 970-979; Saalbach et al. (1994) *Mol. Gen. Genetics* 242, 226-236). Chimeric genes linking the coding regions of 19 and 23 kDa corn storage proteins to Cauliflower Mosaic Virus 35S promoter were found to be expressed at low levels in seeds roots and leaves of transformed tobacco (Scherthaner et al., (1988) *EMBO J.* 7, 1249-1255). Replacement of the monocot regulatory regions with similar regions from dicots resulted in low level seed specific expression of a 19 kDa zein in transformed petunia (Williamson et al., (1988) *Plant Physiol.* 88, 1002-1007) and tobacco (Ohtani et al., (1991) *Plant Mol. Biol.* 16, 117-128). In another case, high level seed-specific expression of the 15 kDa sulfur rich zein was found in transformed tobacco and the signal sequence of the monocot precursor was also correctly processed (Hoffman et al., (1987) *EMBO J.* 6, 3212-3221).

Two bodies of work have previously demonstrated that limited changes can be made to the sequence of 2S albumins which result in stably accumulated modified 2S albumins in plant seeds. European Patent Publication No. EP-0-318,341, Vandekerckhove et al. (1989), *Bio/Technology* 7, 929-932) and De Clercq et al. (1990) *Plant Physiol.* 94, 970-979) demonstrate that changes limited to the region between the

6th and 7th cysteine residues can be made. No changes beyond that region were made, and the authors teach that changes elsewhere in the protein may disrupt the structure because of the conservation of distances between the other cysteine residues. The claims of Ballo (World Patent Publication No. WO 94/10315) have disclosed that replacing arginine by essential lysine of a putative 2S albumin sequence should be tolerated by the protein. Since both amino acids are basic residues, such a change was described as conservative. The details did not disclose whether the protein was able to fold or accumulate in plants. It was also disclosed that the replacements should not create any pairs of adjacent amino acids not found in the natural or homologous seed storage proteins. Because of this latter constraint, not all arginine residues in the putative protein were replaced with lysine. These two works thus teach that 2S albumins can be modified in strictly defined ways, limited either to a particular region or conservative amino acid changes in specific positions.

#### SUMMARY OF THE INVENTION

This invention pertains to a modified Brazil Nut 2S albumin seed storage protein wherein: (i) the amino acid sequence of the modified protein is at least 40% homologous to the wild type Brazil Nut 2S albumin seed storage protein; (ii) all cysteine residues of the modified protein are conserved relative to the wild type protein; (iii) at least 40% of proline residues are conserved relative to the wild type protein; (iv) at least 80% of leucine residues are conserved relative to the wild type protein; and (v) the modified protein comprises at least one non-conservative amino acid substitution not within the hypervariable loop, the substitution consisting of replacement of a non-essential amino acid with an essential amino acid.

A preferred embodiment of the instant invention is a modified Brazil Nut 2S albumin seed storage protein wherein: (i) the amino acid sequence of the modified protein is at least 82% homologous to the wild type Brazil Nut 2S albumin seed storage protein; (ii) all cysteine, proline, leucine and methionine residues of the modified protein are conserved relative to the wild type protein; (iii) all arginine residues of the wild type protein are substituted with lysine residues; and (iv) the modified protein comprises at least three non-conservative amino acid substitutions not within the hypervariable loop, said substitutions comprising substituting two glutamic acid residues with lysine residues and substituting one glutamine residue with a lysine residue.

Another embodiment of the instant invention is an isolated nucleic acid fragment encoding the modified Brazil Nut 2S albumin seed storage protein described above, and a chimeric gene wherein the nucleic acid fragment operably linked to suitable regulatory sequences.

A further embodiment of the instant invention is a transformed plant comprising in its genome the chimeric gene described above. Preferred plants include soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, rice and forage grasses.

Yet another embodiment of the instant invention are seeds derived from the transformed plant described above wherein the seeds comprise the chimeric gene.

Still another embodiment of the instant invention is a method for increasing the essential amino acid content of seeds, the method comprising: (a) preparing a nucleic acid fragment encoding a modified Brazil Nut 2S albumin seed storage protein wherein

(i) the amino acid sequence of the modified protein is at least 40% homologous to the wild type Brazil Nut 2S albumin seed storage protein; (ii) all cysteine residues of the modified protein are conserved relative to the wild type protein; (iii) at least 40% of proline residues are conserved relative to the wild type protein; (iv) at least 80% of leucine residues are conserved relative to the wild type protein; and (v) the modified

protein comprises at least one non-conservative amino acid substitution not within the hypervariable loop, the substitution consisting of replacement of a non-essential amino acid with an essential amino acid; (b) preparing a chimeric gene comprising the nucleic acid fragment of step (a) operably linked to suitable regulatory sequences;

(c) transforming a plant with the chimeric gene of step (b); and (d) obtaining seeds from the transformed plant of step (c).

#### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in *Nucleic Acids Research* 13, 3021-3030 (1985) and in the *Biochemical Journal* 219(2), 345-373 (1984) which are incorporated by reference herein.

Figure 1 is the nucleotide sequence and deduced amino acid sequence of the wild type Brazil Nut 2S albumin gene in plasmid pBNwt that was used as the starting point for the genetic modifications described herein. Relevant restriction enzyme cleavage sites are indicated.

Figure 2 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BNCNS.

Figure 3 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN11.

Figure 4 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN15.

Figure 5 is the nucleotide sequence and deduced amino acid sequence of the  
- modified Brazil Nut 2S albumin gene BN17.

Figure 6 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene BN18.

5        Figure 7 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene BN19.

Figure 8 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene BN153KW.

10       Figure 9 is a composite of the amino acid sequences encoded by the wild type  
Brazil Nut 2S albumin gene (wt) and modified Brazil Nut 2S albumin genes  
exemplified herein. Sulfur-containing amino acid residues are indicated in bold.

Figure 10 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene AT2S1BN15.

15       Figure 11 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene AT2S1BN19.

Figure 12 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene AT2S1BN153W.

20       SEQ ID NO:1 is the nucleotide sequence and deduced amino acid sequence of the  
wild type Brazil Nut 2S albumin gene in plasmid pBNwt that was used as the starting  
point for the genetic modifications described herein.

SEQ ID NO:2 is the amino acid sequence of the wild type Brazil Nut 2S albumin  
protein.

SEQ ID NOs:3-6 are four synthetic oligonucleotides used in the construction of  
the modified Brazil Nut 2S albumin gene BNCNSS.

25       SEQ ID NO:7 is the nucleotide sequence and deduced amino acid sequence of the  
modified Brazil Nut 2S albumin gene BNCNSS.

SEQ ID Nos:8 and 9 are two synthetic oligonucleotide used in the construction of  
the modified Brazil Nut 2S albumin gene BN11.

30       SEQ ID NO:10 is the nucleotide sequence and deduced amino acid sequence of  
the modified Brazil Nut 2S albumin gene BN11.

SEQ ID NOs:11-14 are four synthetic oligonucleotides used in the construction of  
the modified Brazil Nut 2S albumin gene BN15.

SEQ ID NO:15 is the nucleotide sequence and deduced amino acid sequence of  
the modified Brazil Nut 2S albumin gene BN15.

35       SEQ ID Nos:16 and 17 are two synthetic oligonucleotides used in the construction  
of the modified Brazil Nut 2S albumin gene BN17.



SEQ ID NO:18 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN17.

SEQ ID Nos:19 and 20 are two synthetic oligonucleotides used in the construction of the modified Brazil Nut 2S albumin gene BN18.

5 SEQ ID NO:21 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN18.

SEQ ID Nos:22 and 23 are two synthetic oligonucleotides used in the construction of the modified Brazil Nut 2S albumin gene BN19.

10 SEQ ID NO:24 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN19.

SEQ ID Nos:25-28 are four synthetic oligonucleotides used in the construction of the modified Brazil Nut 2S albumin gene BN153KW.

SEQ ID NO:29 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene BN153KW.

15 SEQ ID Nos:30 and 31 are two synthetic oligonucleotides used in the construction of the Brazil Nut 2S albumin genes comprising the Arabidopsis 2S albumin precursor sequence.

SEQ ID NO:32 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene AT2S1BN15.

20 SEQ ID NO:33 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene AT2S1BN19.

SEQ ID Nos:34 and 35 are two synthetic oligonucleotides used in the construction of the modified Brazil Nut 2S albumin gene AT2S1BN153W.

25 SEQ ID NO:36 is the nucleotide sequence and deduced amino acid sequence of the modified Brazil Nut 2S albumin gene AT2S1BN153W.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention demonstrates that the 2S albumin from Brazil Nut is able to accommodate much more radical changes than had been demonstrated previously and that non-conservative replacements with the intent of enriching the protein with essential amino acids outside of the region between the 6th and 7th cysteines can be tolerated without influencing the ability of the protein to be expressed in the seeds of transgenic plants. Such altered Brazil Nut 2S albumins modified such that they are composed of more than two essential amino acids can accumulate to sufficiently high levels to influence the nutritional value of the seed protein.

35 The present invention describes nucleic acid fragments that encode a modified high sulfur 2S albumin seed storage protein. This novel protein is analogous to the protein isolated from Brazil Nuts that is rich in methionine and cysteine amino acids,

but has been altered to include other essential amino acids using site specific replacement techniques.

The structure of the wild type Brazil Nut 2S albumin protein (Figure 1; SEQ ID NO:2) is characterized by the presence of eight cysteine residues that form four disulfide bonds. Twenty percent of the sequence is methionine, distributed throughout the sequence both as isolated residues and in denser regions of adjacent occupancy. Between the 6th and 7th cysteine residues is a region that has been termed the "hypervariable loop," so designated because there are examples of engineered versions of the *Arabidopsis* homolog of the Brazil Nut 2S albumin where substantial amounts or even all of this segment, except for four amino acids immediately adjacent to the amino end of the region and 5 amino acids adjacent to the carboxyl end of the hypervariable loop have been replaced with other non-related sequences.

The Brazil Nut protein is also distinct from other 2S albumins because it is rich in arginine, glutamine and some glutamic acid residues. The presence of a significant fraction (15%) of residues that are basic, as arginine for example, suggested that these positions might be suitable for replacement with lysine, an amino acid that is also basic but unlike arginine, is an essential dietary requirement for animals. Furthermore, it was also considered possible that more radical alterations might be achievable, since protein folding might depend mainly on the correct formation of disulfide bonds rather than the identity of other residues. Genes have been constructed that encode a protein that has all of the native argininy residues replaced by lysyl residues and then further supplemented with additional lysyl residues by alterations at positions not expected to tolerate such changes. These genes have been expressed in a microorganism and in a transgenic plant to alter the nutritional quality of the seed proteins. The increase in methionine and lysine in the seed must be determined by a) the level of expression of the engineered gene in the transformed plant, which depends in part on the seed specific expression signals that are used, b) the percentage of methionine and lysine in the coding region of the engineered gene, c) the stability of the expressed protein in the seed of the transformed plant, which depends in part on its correct processing, intracellular targeting, folding into a structure to allow accumulation in the seed, and ability to withstand desiccation, and d) the compatibility of the new protein with the natural variants of the transformed plant.

Transfer of the gene constructs of the invention (linked to suitable regulatory sequences) into a living cell will result in the production of the encoded protein. Additionally, transfer of the gene constructs of the invention into plants, particularly *Brassica*, or other suitable crop plants such as corn, soybean or oil seed rape, with suitable regulatory sequences to direct expression of the protein in seeds may result in

increased level of sulfur-containing and basic amino acids, particularly methionine and lysine, respectively, thus improving the nutritional quality of seed protein for animals.

#### Definitions

The following terms shall have the meaning set forth herein: The term "essential amino acids" refers to those amino acids which must be obtained by animals and humans from dietary sources. The essential amino acids are arginine (Arg), histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp) and valine (Val).

The term "nucleic acid" refers to a polynucleotide of high molecular weight which can be single-stranded or double-stranded, composed of monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers.

The term "homologous to" refers to the complementarity between the nucleotide sequence of two nucleic acid molecules or between the amino acid sequences of two protein molecules. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (as described in Hames and Higgins (eds.) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.); or by the comparison of sequence similarity between two nucleic acids or proteins.

The term "substantially similar" refers to nucleotide and amino acid sequences that represent equivalents of the instant inventive sequences. For example, altered nucleotide sequences which simply reflect the degeneracy of the genetic code but nonetheless encode amino acid sequences that are identical to the inventive amino acid sequences are substantially similar to the inventive sequences. In addition, amino acid sequences that are substantially similar to the instant sequences are those wherein overall amino acid identity is 95% or greater to the instant sequences. Modifications to the instant invention that result in equivalent nucleotide or amino acid sequences is well within the routine skill in the art. Moreover, the skilled artisan recognizes that equivalent nucleotide sequences encompassed by this invention can also be defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the nucleotide sequences that are within the literal scope of the instant claims.

The term "gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. "Native" gene refers to the gene as found in nature with its own regulatory sequences. "Chimeric" gene refers to a gene comprising heterogeneous regulatory and coding sequences. "Endogenous" gene refers to the native gene normally found in its natural location in the genome. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

The term "coding sequence" refers to a DNA sequence that codes for a specific protein and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The terms "initiation codon" and "termination codon" refer to units of three adjacent nucleotides in a coding sequence that specify initiation and chain termination, respectively, of protein synthesis (mRNA translation). "Open reading frame" refers to the region of a DNA or RNA that is between translation initiation and termination codons and is therefore capable of encoding a protein product.

The term "RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

The term "regulatory sequences" means nucleotide sequences located upstream (5'), within, and/or downstream (3') to a coding sequence, which control the transcription and/or expression of the coding sequences, potentially in conjunction with the protein biosynthetic apparatus of the cell. These nucleotide sequences include a promoter sequence, a translation leader sequence, a transcription termination sequence, and a polyadenylation sequence.

The term "promoter" refers to a DNA sequence in a gene, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. A promoter may also contain DNA sequences that are involved in the binding of protein

factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions. It may also contain enhancer elements.

The term "enhancer" means a DNA sequence which can stimulate promoter activity. It may be an innate element of the promoter or a heterologous element inserted to enhance the level and/or tissue-specificity of a promoter. "Constitutive promoters" refers to those promoters that direct gene expression in substantially all tissues and at substantially all times. "Organ-specific" or "development-specific" promoters as referred to herein are those that direct gene expression almost exclusively in specific organs, such as leaves or seeds, or at specific development stages in an organ, such as in early or late embryogenesis, respectively.

The term "expression" means the production of the protein product encoded by a gene.

The term "3' non-coding sequences" refers to the DNA sequence portion of a gene that contains a transcription termination signal, polyadenylation signal, and any other regulatory signal capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The term "5' non-coding sequences" refers to the DNA sequence portion of a gene that contains a promoter sequence and a translation leader sequence.

The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

The term "mature" protein refers to a post-translationally processed polypeptide without its signal peptide. "Precursor" protein refers to the primary product of translation of an mRNA. "Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide.

The term "intracellular localization sequence" refers to a nucleotide sequence that encodes an intracellular targeting signal. An "intracellular targeting signal" is an amino acid sequence which is translated in conjunction with a protein and directs it to a particular sub-cellular compartment. "Endoplasmic reticulum (ER) stop transit signal" refers to a carboxy-terminal extension of a polypeptide, which is translated in conjunction with the polypeptide and causes a protein that enters the secretory pathway to be retained in the ER. "ER stop transit sequence" refers to a nucleotide sequence that

encodes the ER targeting signal. Other intracellular targeting sequences encode targeting signals active in seeds and/or leaves and vacuolar targeting signals.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms".

Examples of methods of transformation of plants and plant cells include

*Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143, 277) and particle bombardment technology (Klein et al. (1987) *Nature (London)* 327, 70-73; U.S. Patent No. 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known to the skilled artisan (see, for example, Fromm et al. (1990) *Bio/Technology* 8, 833).

The term "cassette" means a nucleic acid fragment prepared by the annealing of two synthetic and complementary oligonucleotides.

Based upon published sequences (Altenbach et al., (1987) *Plant Molecular Biology* 8, 239-250; Gander et al., (1991) *Plant Molecular Biology* 16, 437-448) of the Brazil Nut 2S albumin gene, oligonucleotides were synthesized to allow construction of modified forms of the wild type gene, either through mismatch site-specific procedures or as double-stranded DNA cassettes. The first mutations were introduced into the wild type 2S gene using oligonucleotides that were complementary to the sequence except in those positions coinciding with the places where base changes were desired. The construct used to achieve these changes was an M13-based plasmid that allowed isolation of single stranded form of the wild type 2S albumin gene (pM13BNwt). The use of four oligonucleotides (SEQ ID NOs:3-6) through only one round of mutagenesis produced a version of the gene with 7 of the 15 Arg residues replaced by Lys and also coincidentally introduced new unique restriction sites, SacII, StuI, NheI and ClaI. The final version also had the NcoI site of the wild type gene removed. This version was designated pBNCNSS (Figure 2; SEQ ID NO:7). Although removal of the NcoI site meant loss of one Met residue, this was compensated by the N-terminal Met introduced with the NdeI site at the 5'-end of the gene of the original construct.

Replacement of Arg codons in the gene with Lys codons was achieved in a progressive fashion using the new restriction sites available in pBNCNSS (Figure 2) and replacement of segments of the gene with double-stranded cassettes. It has been our experience construction of genes using large cassettes is not predictably successful and often results in poor efficiency in the number of transformants that carry the desired DNA after amplification. This may be due to the bacterium using correcting mechanisms when transformed with relaxed duplex DNA. The oligonucleotides

synthesized to form the cassettes were therefore routinely extended at the 5' ends so that once the complementary pairs of strands had been annealed to form the cassette, restriction digestion would result in fragments bearing ends of single stranded DNA with precise complementarity to the single stranded ends of a vector that had been treated with the same restriction enzyme; these could then be ligated more efficiently into the expression vector of choice. By introducing more restriction sites or removal of existing sites, the altered DNA from each series of new constructs could be readily assessed for success. All new versions of the gene were sequenced fully to confirm that the desired mutations had been introduced. As protein expression was ultimately to be assessed with progressive enrichment of Lys, this version of the gene was introduced into pET24a (Novagen, Inc., Madison, WI), a commercial plasmid allowing expression from the T7 promoter in suitable hosts. This provided plasmid pETBNCNSS.

The first construct that was made using cassettes, pETBN11, comprising the modified Brazil Nut 2S albumin gene BN11 (Figure 3; SEQ ID NO:10), is characterized by the introduction twelve of Lys codons (compared to zero Lys codons in the wild type 2S gene). In addition, the threonine (Thr) codon corresponding to position 33 of the wild type protein was replaced with serine (Ser) codon to introduce an SphI site adjacent to the SacII site.

Complete replacements for all Arg codons was achieved using two sets of double stranded cassettes (SEQ ID NOs:11-14) to replace the SphI to HindIII segment of BN11. In this case, an internal StyI was created for convenient ligation of the two sets of cassettes after annealing, and the opportunity was taken to remove 125 bases from the 3'-non-coding region of BN11. The resulting construct, pETBN15, was the version of the enriched gene (BN15; Figure 4; SEQ ID NO:15) with all fifteen Arg codons replaced by Lys codons and from which all other variants were made.

Variants of BN15 that explored introduction of other essential amino acids through non-conservative changes included BN17 (Figure 5; SEQ ID NO:18) wherein Ser 107 changed to Lys and glycine (Gly) 105 to Lys by oligonucleotide replacement (SEQ ID NOs:16 and 17). Likewise, BN18 was a version of BN17 wherein Ser 44 changed to Lys (Figure 6; SEQ ID NO: 21) by oligonucleotide replacement (SEQ ID NOs:19 and 20).

A version of BN15 was further enhanced with Lys and Met residues by replacing amino acids that are not considered of this type, i.e., non-conservative changes. Thus replacement of the NdeI-SphI fragment of BN15 with two oligonucleotides (SEQ ID NOs:22 and 23) introduced two further Lys residues in place of two glutamic acid (Glu) residues at positions 4 and 28, and introduced an extra Met replacing Glu 27. This produced the modified Brazil Nut 2S albumin gene BN19 (Figure 7; SEQ ID NO:24).

Other variants such as BN153KW (Figure 8; SEQ ID NO:29), explored introduction of the essential amino acid Trp into the sequence of the Brazil Nut 2S albumin. Figure 9 is a composite of the amino acid sequences encoded by each of the wild type and modified Brazil Nut 2S albumin genes described above.

5 In order to create the gene that encodes the precursor form of the 2S storage protein, nucleic acid fragments were constructed that recreated the precursor sequence of an *Arabidopsis* 2S protein. An oligonucleotide cassette (comprising SEQ ID NOs:30 and 31) encoding the *Arabidopsis* precursor sequence was designed so that introduction of the cassette into the 5'-end of the wild type and modified 2S albumin genes described above resulted in an in-frame fusion of the precursor sequence with the sequence of the mature genes. The resulting genes had an NcoI site at the ATG initiation codon of precursor sequence and an NdeI site at the first codon of the mature sequence. These precursor genes were designated AT2S1BNwt (not shown), AT2S1BN15 (Figure 10; SEQ ID NO:32), and AT2S1BN19 (Figure 11; SEQ ID NO:33) to indicate they contained the wild type and BN15, and BN19 variants of the 2S albumin gene, respectively. Another precursor gene, AT2S1BN153W (Figure 12; SEQ ID NO:36), was prepared by direct replacement of the NheI to HindIII fragment of pAT2S1BN15 with a cassette which directed the incorporation of an increased number of tryptophanyl residues in the gene product.

20 The nucleic acid fragment coding for the sulfur rich seed 2S protein may be attached to suitable regulatory sequences and used to overproduce the protein in microbes such as bacteria or yeast, or in transgenic plants such as *Brassica*, cereals or legumes. Such a DNA construction may include either the wild type 2S gene or an engineered gene. One skilled in the art can isolate the coding sequences from the fragment of the invention by using and/or creating restriction endonuclease sites.

#### Expression of enriched 2S protein in *E. coli*

To express the modified Brazil Nut 2S coding sequences in *E. coli*, the commercial expression vector pET24a was used. This vector employs the bacteriophage T7 RNA polymerase/T7 promoter system (Studier et al., (1990) *Methods in Enzymology* 185, 60-89) for gene transcription. The variants of all the 2S albumin genes including the wild type construct were ligated into pET24a using the NdeI-HindIII sites. These constructs were used to transform competent *E. coli* cells (strain BL21) which were grown to mid-log phase in LB before induction with IPTG. The protein expressed in transformed *E. coli* hosts, was assessed by electrophoresis of lysed cell contents on SDS polyacrylamide gels and comparison to authentic Brazil Nut 2S albumin isolated from the native source. Verification that the expressed protein bands on gels included the recombinant 2S proteins was achieved by electroblotting the



proteins to PVDF membranes, obtaining the N-terminal sequences, and confirming that the experimental sequences matched the predicted sequences.

Expression of enriched 2S protein in plants

The nucleic acid fragments of the invention can be used to produce large quantities of the 2S protein enriched in essential amino acids especially methionine and lysine via fermentation in *E. coli* or other microorganisms. To do this the nucleic acid fragment of the invention can be operably linked to a suitable regulatory sequence comprising a promoter sequence, a translation leader sequence and a 3' non-coding sequence. The chimeric gene can then be introduced into a microorganism via transformation and the transformed organism grown under conditions resulting in high expression of the engineered gene. The cells containing the protein rich in essential amino acids can be collected and the enriched protein extracted. Because high level production is not toxic to the cells, higher levels could be achieved using other strains.

A preferred class of hosts for the expression of the coding sequence of modified Brazil Nut 2S albumin proteins are eukaryotic hosts, particularly the cells of higher plants. Particularly preferred among the higher plants and the seeds derived from them are soybean, rapeseed (*Brassica napus*, *B. campestris*), sunflower (*Helianthus annuus*), cotton (*Gossypium hirsutum*), corn, tobacco (*Nicotiana tabacum*), alfalfa (*Medicago sativa*), wheat (*Triticum sp.*), barley (*Hordeum vulgare*), oats (*Avena sativa*, L), sorghum (*Sorghum bicolor*), rice (*Oryza sativa*), and forage grasses. Expression in plants will use regulatory sequences functional in such plants.

The expression of foreign genes in plants is well-established (De Blaere et al.(1987) *Meth. Enzymol.* 153:277-291). The origin of promoter chosen to drive the expression of the coding sequence is not critical as long as it has sufficient transcriptional activity to accomplish the invention by increasing the level of translatable mRNA for modified Brazil Nut 2S albumin proteins in the desired host tissue. Preferred promoters for expression in all plant organs, and especially for expression in leaves include those directing the 19S and 35S transcripts in Cauliflower mosaic virus (Odell et al. (1985) *Nature* 313, 810-812; Hull et al. (1987) *Virology* 86, 482-493), small subunit of ribulose 1,5-bisphosphate carboxylase (Morelli et al.(1985) *Nature* 315, 200; Broglie et al. (1984) *Science* 224, 838; Herrera-Estrella et al.(1984) *Nature* 310, 115; Coruzzi et al.(1984) *EMBO J.* 3, 1671; Facciotti et al.(1985) *Bio/Technology* 3, 241), maize zein protein (Matzke et al.(1984) *EMBO J.* 3, 1525), and chlorophyll a/b binding protein (Lampa et al.(1986) *Nature* 316, 750-752).

Depending upon the application, it may be desirable to select promoters that are specific for expression in one or more organs of the plant. Examples include the light-

inducible promoters of the small subunit of ribulose 1,5-bisphosphate carboxylase, if the expression is desired in photosynthetic organs, or promoters active specifically in seeds.

Preferred promoters are those that allow expression of the protein specifically in seeds. This may be especially useful, since seeds are the primary source of vegetable protein and also since seed-specific expression will avoid any potential deleterious effect in non-seed organs. Examples of seed-specific promoters include, but are not limited to, the promoters of seed storage proteins, which represent more than 50% of total seed protein in many plants. The seed storage proteins are strictly regulated, being expressed almost exclusively in seeds in a highly organ-specific and stage-specific manner (Higgins et al. (1984) *Ann. Rev. Plant Physiol.* 35, 191-221; Goldberg et al. (1989) *Cell* 56, 149-160; Thompson et al. (1989) *BioEssays* 10, 108-113). Moreover, different seed storage proteins may be expressed at different stages of seed development.

There are currently numerous examples for seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean  $\beta$ -phaseolin (Sengupta-Gopalan et al. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3320-3324; Hoffman et al. (1988) *Plant Mol. Biol.* 11, 717-729), bean lectin (Voelker et al. (1987) *EMBO J.* 6, 3571-3577), soybean lectin (Okamuro et al. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8240-8244), soybean kunitz trypsin inhibitor (Perez-Grau et al. (1989) *Plant Cell* 1:095-1109), soybean  $\beta$ -conglycinin (Beachy et al. (1985) *EMBO J.* 4, 3047-3053; Barker et al. (1988) *Proc. Natl. Acad. Sci. USA* 85, 458-462; Chen et al. (1988) *EMBO J.* 7, 297-302; Chen et al. (1989) *Dev. Genet.* 10, 112-122; Naito et al. (1988) *Plant Mol. Biol.* 11, 109-123), pea vicilin (Higgins et al. (1988) *Plant Mol. Biol.* 11, 683-695), pea convicilin (Newbigin et al. (1990) *Planta* 180461), pea legumin (Shirsat et al. (1989) *Mol. Gen. Genetics* 215, 326); rapeseed napin (Radke et al. (1988) *Theor. Appl. Genet.* 75, 685-694) as well as genes from monocotyledonous plants such as for maize 15 kD zein (Hoffman et al. (1987) *EMBO J.* 6, 3213-3221; Schernthaner et al. (1988) *EMBO J.* 7, 1249-1253; Williamson et al. (1988) *Plant Physiol.* 88, 1002-1007), barley  $\beta$ -hordein (Marris et al. (1988) *Plant Mol. Biol.* 10, 359-366) and wheat glutenin (Colot et al. (1987) *EMBO J.* 6, 3559-3564). Moreover, promoters of seed-specific genes operably linked to heterologous coding sequences in chimeric gene constructs also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include *Arabidopsis thaliana* 2S seed storage protein gene promoter to express enkephalin peptides in *Arabidopsis* and *B. napus* seeds (Vandekerckhove et al. (1989) *BioTechnology* 7, 929-932), bean lectin and bean  $\beta$ -phaseolin promoters to express luciferase (Riggs et al.

(1989) *Plant Sci.* 63, 47-57), and wheat glutenin promoters to express chloramphenicol acetyltransferase (Colot et al. (1987) *EMBO J.* 6, 3559-3564).

Of particular use in the expression of the nucleic acid fragment of the invention will be the heterologous promoters from several extensively-characterized soybean seed storage protein genes such as those for the Kunitz trypsin inhibitor (Jofuku et al. (1989) *Plant Cell* 1, 1079-1093; Perez-Grau et al. (1989) *Plant Cell* 1, 1095-1109), glycinin (Nielson et al. (1989) *Plant Cell* 1, 313-328),  $\beta$ -conglycinin (Harada et al. (1989) *Plant Cell* 1, 415-425). Promoters of genes for  $\alpha'$ - and  $\beta$ -subunits of soybean  $\beta$ -conglycinin storage protein will be particularly useful in expressing the modified Brazil Nut 2S albumin 2S albumin mRNA in the cotyledons at mid- to late-stages of soybean seed development (Beachy et al. (1985) *EMBO J.* 4, 3047-3053; Barker et al. (1988) *Proc. Natl. Acad. Sci. USA* 85, 458-462; Chen et al. (1988) *EMBO J.* 7, 297-302; Chen et al. (1989) *Dev. Genet.* 10, 112-122; Naito et al. (1988) *Plant Mol. Biol.* 11, 109-123) in transgenic plants, since: a) there is very little position effect on their expression in transgenic seeds, and b) the two promoters show different temporal regulation: the promoter for the  $\alpha'$ -subunit gene is expressed a few days before that for the  $\beta$ -subunit gene.

Also of particular use in the expression of the nucleic acid fragments of the invention will be the heterologous promoters from several extensively characterized corn seed storage protein genes such as those from the 10 kD zein (Kiriwara et al. (1988) *Gene* 71, 359-370), the 27 kD zein (Prat et al. (1987) *Gene* 52, 51-49; Gallardo et al. (1988) *Plant Sci.* 54, 211-281), and the 19 kD zein (Marks et al. (1985) *J. Biol. Chem.* 260, 16451-16459). The relative transcriptional activities of these promoters in corn have been reported (Kodrzycki et al. (1989) *Plant Cell* 1, 105-114) providing a basis for choosing a promoter for use in chimeric gene constructs for corn or other monocots.

Proper level of expression of 2S engineered genes enriched in essential amino acids may require the use of different promoters. Such chimeras can be transferred into host plants either together in a single expression vector or sequentially using more than one vector or more than one copy of the enriched gene transcribed from the same vector.

It is envisioned that the introduction of enhancers or enhancer-like elements into promoter constructs will also provide increased levels of primary transcription for modified Brazil Nut 2S albumin proteins to accomplish the invention. This would include viral enhancers such as that found in the 35S promoter (Odell et al. (1988) *Plant Mol. Biol.* 10, 263-272), enhancers from the opine genes (Fromm et al. (1989) *Plant Cell* 1, 977-984), or enhancers from any other source that result in increased

transcription when placed into a promoter operably linked to the nucleic acid fragment of the invention.

Of particular importance is the DNA sequence element isolated from the gene for the  $\alpha'$ -subunit of  $\beta$ -conglycinin that can confer 40-fold seed-specific enhancement to a constitutive promoter (Chen et al. (1988) *EMBO J.* 7, 297-302; Chen et al. (1989) *Dev. Genet.* 10, 112-122). One skilled in the art can readily isolate this element and insert it within the promoter region of any gene in order to obtain seed-specific enhanced expression with the promoter in transgenic plants. Insertion of such an element in any seed-specific gene that is expressed at different times than the  $\beta$ -conglycinin gene will result in expression in transgenic plants for a longer period during seed development.

The invention can also be accomplished by a variety of other methods to obtain the desired end. In one form the invention is based on modifying plants to produce increased levels of 2S enriched protein by having significantly larger numbers of copies of the modified gene either through enhanced promotion or multiple copies on each message.

Any 3' non-coding region capable of providing a transcription termination signal, a polyadenylation signal and other regulatory sequences that may be required for the proper expression of the modified Brazil Nut 2S albumin protein coding region can be used to accomplish the invention. This would include the 3' end from a heterologous zein gene, the 3' end from any storage protein such as the 3' end of the soybean  $\beta$ -conglycinin gene, the 3' end from viral genes such as the 3' end of the 35S or the 19S cauliflower mosaic virus transcripts, the 3' end from the opine synthesis genes, the 3' ends of ribulose 1,5-bisphosphate carboxylase or chlorophyll a/b binding protein, or 3' end sequences from any source such that the sequence employed provides the necessary regulatory information within its nucleic acid sequence to result in the proper expression of the promoter/modified Brazil Nut 2S albumin protein coding region combination to which it is operably linked. There are numerous examples in the art that teach the usefulness of different 3' non-coding regions (for example, see Ingelbrecht et al. (1989) *Plant Cell* 1, 671-680).

DNA sequences coding for intracellular localization sequences may be added to the modified Brazil Nut 2S albumin protein coding sequence if required for the proper expression of the proteins to accomplish the invention. Thus the signal sequence from the  $\beta$  subunit of phaseolin from the bean *Phaseolus vulgaris*, or the signal sequence from the  $\alpha'$  subunit of  $\beta$ -conglycinin from soybean (Doyle et al. (1986) *J. Biol. Chem.* 261, 9228-9238), can be employed. Hoffman et al. ((1987) *EMBO J.* 6, 3213-3221) showed that the signal sequence of the moncot precursor of a 15 kD zein directed the protein into the secretory pathway and was also correctly processed in transgenic

tobacco seeds. However, the protein did not remain within the endoplasmic reticulum as is the case in corn. To retain the protein in the endoplasmic reticulum it may be necessary to add stop transit sequences. It is known in the art that the addition of DNA sequences coding for the amino acid sequence (Lys-Asp-Glu-Leu) at the carboxyl terminal of the protein retains proteins in the lumen of the endoplasmic reticulum (Munro et al. (1987) *Cell* 48, 899-907; Pelham (1988) *EMBO J.* 7, 913-918; Pelham et al. (1988) *EMBO J.* 7, 1757-1762; Inohara et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 3564-3568; Hesse et al. (1989) *EMBO J.* 8, 2453-2461). In some plants seed storage proteins are located in the vacuoles of the cell. In order to accomplish the invention it may be necessary to direct the modified Brazil Nut 2S albumin protein to the vacuole of these plants by adding a vacuolar targeting sequence. A short amino acid domain that serves as a vacuolar targeting sequence has been identified from bean phytohemagglutinin which accumulates in protein storage vacuoles of cotyledons (Tague et al. (1990) *Plant Cell* 2, 533-546). In another report a carboxyl-terminal amino acid sequence necessary for directing barley lectin to vacuoles in transgenic tobacco was described (Bednarek et al. (1990) *Plant Cell* 2, 1145-1155).

#### Construction of chimeric genes for expression of Brazil Nut 2S in plants

Three specific gene expression cassettes were used for construction of chimeric genes for expression of 2S in plants to explore expression of altered forms of the gene in a plant host. Specifically those variants of the 2S gene with conservative replacements as exemplified by Arg to Lys and also an example of non-conservative changes as in BN19. The expression cassettes contained the regulatory regions from two highly expressed seed storage protein genes:

- 1) the promoter of the highly expressed storage protein,  $\beta$ -conglycinin of soy bean; and
- 2) the 3'-termination sequence of phaseolin from *Phaseolus vulgaris*.

The precursor sequence of one of the 2S albumin genes from *Arabidopsis thaliana* was introduced in-frame at the 5'-end of the 2S native gene and selected variants to give AT2S1BNwt, AT2S1BN15, AT2S1BN19 and AT2S1BN153W. The precursor versions of these genes were then ligated between the  $\beta$ -conglycinin promoter and the 3'-phaseolin termination region (Slightom et al., (1991) *Plant Mol. Biol. Man.* B16, 1-55) in plasmid pCW109. The vector pCW109 was made by inserting into the HindIII site of the cloning vector pUC18 a 555 bp 5' non-coding region (containing the promoter region) of the  $\beta$ -conglycinin gene followed by the multiple cloning sequence containing the restriction endonuclease sites for Nco I, Sma I, Kpn I and Xba I, then 1174 bp of the common bean phaseolin 3' untranslated region into the HindIII site (described above). This plasmid allows the precursor, mature and flanking regulatory regions to be isolated

- as one large HindIII fragment after amplification and isolation from *E. coli* (Odell et al., (1994) *Plant Physiol.* 106, 447-458). Introduction of the HindIII fragment into the same site of pZS96 (Odell et al., (1994) *Plant Physiol.* 106, 447-458) positions the segment conveniently between the left and right border DNA sequences of the
- 5 Ti plasmid of *Agrobacterium tumefaciens* effective in infecting plant hosts.

- Various methods of transforming cells of higher plants according to the present invention are available to those skilled in the art. A method that found particular use in this case was the infection of *Arabidopsis* plants by vacuum infiltration as described by Bechtold et al. ((1993) *C.R. Acad. Sci. Paris* 316, 1194-1199). Seeds of T3 and T4
- 10 generations of *Arabidopsis* plants harboring the Brazil Nut 2S wild type albumin gene and enhanced variants were studied for altered amino acid enrichment of Lys and Met residues to show that significant increases in the amounts of these two essential amino acids are detectable.

### EXAMPLES

- 15 The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this
- 20 invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

- Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor
- 25 Laboratory Press: Cold Spring Harbor, 1989.

### EXAMPLE 1

#### Molecular Cloning of the Brazil Nut 2S gene

- Genes and cDNAs encoding the Brazil nut 2S albumin have been extensively described (Altenbach et. al., (1987) *Plant Mol. Biol.* 8, 239-250; Gander et. al., (1991)
- 30 *Plant Mol. Biol.* 16, 437-448) and have even been constructed *de novo* on the basis of published sequences (Saalbach et. al. (1994) *Mol. Gen. Genet.* 242, 226-236). The 333bp encoding the mature sequence is shown in Fig. 1, SEQ ID NO:1. The starting Brazil nut 2S albumin sequence (SEQ ID No:1) used herein is supplemented with an initiation codon to facilitate expression in prokaryotic cells. An NdeI-EcoRI fragment
- 35 encompassing the nature sequence was ligated into a derivative of pET3a (Novagen Inc., Madison, WI) that has the NcoI site replaced by a short multiple cloning site containing Nde I and EcoRI. This plasmid, pET3am was digested with NdeI and EcoRI

to accept the BNwt gene fragment giving the plasmid pBNwt; see Figure 3. This plasmid, which also carries the gene for  $\beta$ -lactamase, was used to transform competent *E. coli* (JM 83) cells that were grown in SOC medium (Hanahan, D. (1983) *J. Mol. Biol.* 166, 557) before selecting for plasmid-bearing organisms with ampicillin. The cells were streaked onto agarose-LB plates that contained ampicillin (50 ug/mL) and grown overnight at 37°. A single colony was picked and inoculated into 50 mL of LB medium also containing ampicillin. The culture was shaken at 37° until the cells had reached an OD<sub>600</sub> of about 3.0. The cells were harvested by centrifugation and the DNA isolated and purified using the procedures described by the suppliers of the Promega Wizard™ kit. The purified DNA was verified by restriction site digestion and electrophoretic separation of the fragments on 1% agarose gels. The 2S gene in the plasmid was also sequenced in both directions using primers that annealed to the vector sequence close to the T7 promoter outside the 5' end of the coding region and the 3' end at the T7 terminator.

## EXAMPLE 2

### Modification of the Brazil Nut 2S Albumin Gene by Mutagenesis of Single-Stranded DNA

The 2S gene was excised from pBNwt using unique EcoRI and XbaI sites. The fragment was ligated into M13mp18 that had also been cut with EcoRI and XbaI. The resulting plasmid, pM13BNwt, thus allowed either double- or single-stranded forms of the gene to be isolated. The first series of mutations that resulted in conversion of Arg codons to Lys codons were achieved using the single stranded form of pM13BNwt. The Muta-Gene™ kit (Biorad Laboratories, Richmond, CA) provides a means of strongly selecting against the non-mutagenised strand of double-stranded DNA. This was achieved by transforming *E. coli* CJ236 competent cells with pM13BNwt, a host that has a double mutation in the *dut* and *ung* genes. Some of the thymines in the DNA are stably replaced by uracil in this double mutant. The transformants were grown on LB plates containing chloramphenicol. One of the colonies was grown overnight in chloramphenicol-containing medium and the single-stranded DNA containing uracil was then isolated as a phagemid as described by the suppliers of the kit.

The four oligonucleotides used to introduce mutations into the 2S gene (SEQ ID NOs:3-6) were first phosphorylated with T4 polynucleotide kinase at 37° for 60 min and the reaction stopped by heating at 65° for 10 min. The oligonucleotides were simultaneously annealed to the uracil enriched single-stranded DNA of pM13BNwt at 70° followed by slow cooling to room temperature over 40 min. The resulting partial duplex was then stored on ice before the double-stranded DNA was generated using T4 polymerase to extend the oligonucleotides in the presence of all four dNTPs and T4

DNA ligase to ligate the ends of the extended DNA. The resulting double-stranded DNA was used to transform *E. coli* MV1190 cells that have active uracil-N-glycosylase that inactivates the uracil-containing strand so that only the mutant strand replicates. The transformed cells generated the double stranded form of the M13 derivative; this was assessed by restriction analysis to ensure that the four unique restriction sites SacII, StuI, Nhe I and Cla I engineered into the oligonucleotides had been incorporated as a result of the manipulations, and that the internal NcoI site of the wild type gene was eliminated. The modified gene was excised from the M13 construct using EcoRI and XbaI and ligated back into pET3am. The resulting plasmid, pBNCSS, contains a Brazil Nut 2S albumin gene with seven Arg codons replaced by Lys codon at positions corresponding to amino acids 37, 58, 63, 82, 83, 86, and 101 of the wild type protein, accompanied by a Met to phenylalanine (Phe) at position 104 (Figure 2; SEQ ID NO:7).

### EXAMPLE 3

#### Cassette Mutagenesis of the Brazil Nut 2S Albumin Gene

The modified 2S gene from pBNCSS was isolated by restriction enzyme digestion with NdeI and HindIII, and ligated into pET24a to give pETBNCSS. The series of mutations that replaced a further four of the Arg residues with Lys were localized in the N-terminal half of the gene and were accomplished by annealing synthetic complementary oligonucleotides that coded for the altered sequence from the NdeI site to SacII site (SEQ ID NOs:8 and 9). The individual 131 base oligonucleotides were first purified by electrophoresis on 8% polyacrylamide gel. The band was excised from the gel, eluted, and washed prior to annealing at 90° for 3 min. The annealing solution was then cooled slowly to 30° and placed on ice for 3 min. The oligonucleotides were designed with extended ends beyond the Nde I and Sac II sites so that, following annealing, the double-stranded cassette could be digested with these two enzymes to produce a high percentage of clean 'restriction' ends. The resulting efficiencies and consistency of ligation into the Nde I/Sac II-digested pETBNCSS vector with this cassette approach was evident from the number of transformants carrying the synthetic oligonucleotide insert. The isolated vector was validated with respect to the correct insertion of the cassette by restriction analysis to show the presence of the new Sph I site introduced with the insert and by sequencing the region of the gene coding for the N-terminal segment of the protein. The resulting construct, pETBN11, comprising the BN11 gene, contained twelve Lys codons (Figure 3; SEQ ID NO:10).

The complete replacement of all codons for Arg residues with ones for Lys in the gene was accomplished in a similar fashion but using an oligonucleotide cassette designed to replace the portion of the gene encoding the C-terminal half of the protein. This region of the gene is readily replaced using the convenient Sph I and Hind III sites



in the middle and 3'-end of the gene, respectively. The opportunity was also taken to remove much of the non-coding 3'-end of the original constructs. The number of bases this region covers is too long to be substituted by a cassette formed from only two oligonucleotides. Accordingly, four oligonucleotides were designed that would ultimately be ligated together before insertion into the appropriately cut vector. The first half of the cassette using the two oligonucleotides displayed as SEQ ID NOs: 11 and 14 had an SphI site located 8 bases from the 5'-end and a StyI site located 8 bases from the 3'-end. The second half of the cassette (SEQ ID NOs: 12 and 13) was arranged with a StyI site 8 bases from the 5'-end and HindIII site located some 5 bases from the 3' end. Once each pair of the two halves had been independently annealed, they were digested with StyI, and the StyI ends ligated. The resulting ligated double cassette was isolated from an 8% polyacrylamide gel, washed and digested with SphI and HindIII before ligating into pETBN11 that had previously digested with the SphI and HindIII and isolated from a 1% agarose gel.

Transformants of competent *E. coli* carrying the altered vector were isolated and the DNA purified. The DNA was validated by restriction analysis before sequencing the appropriate region of the gene. In this case, the introduction of a StyI site and removal of SacII was diagnostic of successful construction. The derivative of pETBN11, now with all fifteen Arg replaced by Lys, was designated pETBN15 and encoded the modified BN15 gene (Figure 4, SEQ ID NO:15).

#### EXAMPLE 4

##### Further Enhanced Mutations of the Brazil Nut 2S Albumin Gene

Other mutations were introduced into the gene to explore whether other residues with non-basic sidechains might be sites that are suitable for replacing with essential amino acids. The plasmid used to do these changes was pETBN15. Two residues chosen for replacement with Lys were Gly 105 and Ser 107. The relevant region of the gene has NheI and BamHI restriction sites that are convenient for the purpose of introducing substitutions. The 85 bp NheI-BamHI segment was replaced in pETBN15 with a cassette synthesized with the NheI and BamHI sites indented by 9 bases (SEQ ID NOs:16 and 17). The cassette was first digested with the two restriction enzymes to provide clean Nhe I and BamHI ends and the fragment purified by gel electrophoresis. The purified fragment was ligated into pETBN15 that had been cut with the same enzymes. The resulting plasmid was termed pETBN17 and encoded the modified Brazil Nut 2S albumin gene designated BN17 (Figure 5; SEQ ID NO:18).

Using similar procedures, the serine (Ser) residue at position 44 was replaced by Lys in pETBN17 using a cassette formed by annealing the oligonucleotides shown in SEQ ID NOs:19 and 20 and replacement of the SphI-StuI segment of the gene BN17.

The resulting construct designated pETBN18 and encoded the modified Brazil Nut 2S albumin gene BN18 (Figure 6; SEQ ID NO:21).

Replacement of the NdeI-SphI fragment of pETBN15 with a cassette formed by annealing the oligonucleotides depicted in SEQ ID NOs:22 and 23 resulted in alterations of Glu 4 and 28 to Lys and Glu 27 to Met. This construct was designated pETBN19 and encoded the BN19 mutant (Figure 7; SEQ ID NO:24). Replacement of NdeI-StuI fragment of pETBN15 with two ligated cassettes formed by annealing the oligonucleotides depicted in SEQ ID NOs:25 and 26 and oligonucleotides depicted in SEQ ID Nos:27 and 28 introduced three further Lys at positions 4, 16 and 41, replacing Glu, Ser and Pro, respectively; and one Trp at position 42 replacing His, to give pETBN153KW, encoding the modified gene BN153KW (Figure 8; SEQ ID NO:29). The base changes that resulted in Ser to Lys also introduced an AflIII site into the gene, and the opportunity to use a silent mutation changing base 150 from G to C introduced an XhoI site.

#### EXAMPLE 5

##### Construction of a Precursor Form of the Brazil Nut 2S Albumin Genes

The 5'-end of the Brazil Nut 2S albumin gene was extended in-frame to include the 37 amino acid precursor sequence of the Arabidopsis 2S albumin protein that should contain all the information for the correct processing and targeting of the protein in the plant. Two 137 base oligonucleotides were synthesized (SEQ ID NOs:30 and 31) with recessed NcoI and NdeI sites. The fragments were purified and isolated as described above before annealing together to form the double stranded cassette. The cassette was also purified and isolated from an 8% polyacrylamide gel before restriction digestion with NcoI and NdeI to produce the clean 5'- and 3'-ends.

The pETBN15 vector was cut with NdeI and Hind III, and the fragment containing the BN15 gene was purified from the remaining vector using polyacrylamide gels and subsequent elution. The NdeI sites of the cassette and the BN15 gene were then ligated together to produce the extended gene sequence with NcoI and HindIII sites at the 5'- and 3'- ends, respectively. The extended gene was then ligated into pET24d (Novagen Inc., Madison, WI) previously linearized by NcoI and HindIII digestion. The resulting vector was designated pAT2S1BN15 and contains the AT2S1BN15 gene (Figure 10; SEQ ID NO:32).

This construct was then used as the vector into which all versions of the Brazil Nut 2S albumin gene could be inserted to give the 5'- extended gene. For example, construction of the wild type version of the precursor-containing gene was accomplished by replacing the NdeI-HindIII fragment of pAT2S1BN15 with the NdeI-HindIII fragment of pETBNwt that has the wild type sequence of the gene.

Likewise the replacement of the NdeI-HindIII fragment with that of pETBN19

- generated pAT2S1BN19, containing the modified AT2S1BN19 gene (Figure 11; SEQ ID NO:33).

- 5 Smaller segments of the vector could also be manipulated to enhance the amino acid content of the protein products. For example, the NheI to HindIII fragment of pAT2S1BN15 was replaced with a cassette composed of the oligonucleotides depicted in SEQ ID Nos: 22 and 23. This replaced Glu 89, Ser 93 and Phe 104 with Trp giving pAT2S1BN153W (Figure 12; SEQ ID NO:36).

#### EXAMPLE 6

##### 10 Construction of the Plant-Specific Expression Cassette

- The wild type Brazil Nut 2S albumin gene and the modified genes BN15 and BN19 were positioned between the promoter that is normally responsible for controlling conglycinin expression and the 3' region normally found downstream of the phaseolin gene. The vector that contained these control elements (pCW109) was cut with NcoI and SmaI. The plasmids containing the wild type and modified Brazil Nut 2S albumin genes (pAT2S1BNwt, pAT2S1BN15 and pAT2S1BN19) were first cut with EcoRI and blunt-ended with mung bean nuclease. The DNA was precipitated from a solution containing 0.01% SDS and 0.1 M NaCl using two volumes of cold, dry ethanol. The gene encoding the 2S precursor-containing protein was excised from the resulting linearised DNA using NcoI. The NcoI (5'-) blunt-ended (3') fragments from pAT2S1BN15 and pAT2S1BN19 were then ligated into pCW109 that was previously linearised by digestion with NcoI-Sma I. The resulting constructs were designated pCW109BN15 and pCW109BN19. The equivalent version that included the BNwt gene was a little more involved since this gene still has an internal NcoI site. In this case a partial digest of pAT2S1BNwt with NcoI after nuclease treatment of the EcoRI site readily provided the fragment of interest that encompassed the complete gene for the precursor protein that could also be ligated into NcoI-Sma I-digested pCW109. The resulting vector was designated pCW109BNwt.

#### EXAMPLE 7

##### 30 Construction of the Binary Vector Useful for Plant Transformation

- The construction of a vector suitable for plant transformation requires the presence of the right and left border sequences that Agrobacterium utilizes to introduce foreign DNA into the nuclear genome of plants, and also encompasses a selection cassette that allows for antibiotic selection of those plants that show successful integration of the foreign DNA into their genome. Ideally a second selection cassette should also be available for selecting those bacteria transformed with the binary vector for manipulation or amplification.

- The vector chosen to achieve all these desired features was pzs96. pzs96 has genes encoding the N-phosphotransferase that imparts kanamycin resistance on transformed plants and the  $\beta$ -lactamase that imparts ampicillin resistance for selection in bacteria. Each of the three pCW109AT2S1BN plasmids were digested with HindIII.
- 5 The liberated HindIII fragments were purified and then mixed separately with an equimolar amount of HindIII-linearized pzs96 before ligation. After ligation, the DNA was used to transform competent *E. coli*; transformants were selected on media containing ampicillin. Correct construction was assessed by restriction digest analysis and DNA sequencing. In this way only those binary vectors with the Brazil Nut 2S
- 10 albumin genes in the correct orientation were retained for use in transforming plants.

### EXAMPLE 8

#### Transformation of *Arabidopsis thaliana*

- The vacuum infiltration methods described in detail by Bechtold et al. ((1993) *C.R. Acad. Sci. Paris* 316, 1194-1199) were used to infect *Arabidopsis thaliana*
- 15 (ecotype Wassilewskija) with the binary vectors described above. Five to ten plants were grown for 3-5 weeks in pots covered by a fine nylon screen stretched across the top of the pot at the time of seeding to prevent loss of soil. A suspension of *Agrobacterium* that had previously been grown overnight in LB containing kanamycin (25 ug/mL), rifampicin (50 ug/mL) and carbenicillin (100 ug/mL) was dispersed in
- 20 1 liter of infiltration medium to give an OD<sub>600</sub> of 0.8. The bacterial suspension was poured into a tray that was placed into the bottom of the vacuum cabinet. The pots were suspended inverted in the vacuum cabinet and the shoots of the plants submerged in the solution. The door was closed and the vacuum of a rotary vane oil pump applied for 5 min reaching a final vacuum of 1.5-2.0 Torr. The plants (T1 generation) were
- 25 removed and allowed to grow normally and set seed (4 weeks).

- The T2 seeds from this T1 generation were harvested and selected for kanamycin resistance. This selection entailed sterilization of T2 seeds in 50% commercial bleach with 0.02% Tween-20 for 8-10 minutes and then washing in sterile water 3-5 times before sowing. Sterilized seeds were germinated in Petri plates with sterile media
- 30 containing 1/2 strength Murashige-Skoog salts (Gibco #11117-066) plus 0.7% agar, 1% sucrose and 50 ug/mL kanamycin. Kanamycin was prepared as a 50 mg/mL stock in water, sterilized by passage through a 0.2 um filter, and added to the media after it had been autoclaved and cooled to 60°. Plates containing kanamycin were stored at 4° and used within one month of preparation. Those plants that had been successfully
- 35 infected by *Agrobacterium* containing the selection cassette which encompasses the Brazil Nut 2S albumin constructs grew in preference to those without integrated DNA. The plants that survived this selection were transferred from the Petri plates to pots

containing commercial soil mixes (Metromix™ or others) at 1-3 weeks of age.

- Following transplanting, the pots were covered with clear plastic wrap for 3-7 days to allow the seedlings to adapt to the soil conditions. The plastic was then removed and plants were grown to maturity using standard practices in growth chambers at 20-25°
- 5 with fluorescent and incandescent illumination of 100-300  $\mu\text{mol}/\text{m}^2/\text{sec}$  photosynthetically active radiation and a photoperiod ranging from 12 h to continuous illumination. The T2 plants in soil were allowed to self-fertilize to produce the T3 seeds which were harvested and used for analysis.

- To ensure successful integration of the Brazil Nut 2S albumin gene into the
- 10 nuclear genome of this selected group of individual plants, total DNA from the leaves was isolated and PCR used to amplify the Brazil Nut 2S albumin genes and verify that they had also been integrated along with the kanamycin gene. The DNA fragment that resulted from the PCR reactions was further analyzed and confirmed by restriction analysis. The results were compared to the individual examples of the host plant that
- 15 had not been subjected to infection. Once the plants had fully matured, the DNA from seeds was likewise analyzed.

#### EXAMPLE 9

##### Expression of the Brazil Nut 2S Albumin Gene in Transformed Plants

- The seeds from different lines of T2 generation transgenics harboring the
- 20 AT2S1BN15, AT2S1BN19 and AT2S1BNwt genes were harvested for analysis. The seeds (10 mg) from the mature plants were first ground to a fine powder in liquid nitrogen and then defatted at room temp with three washes of *n*-hexane. The resulting defatted flour was allowed to dry before extraction with a weak acidic buffer (0.1M citrate, pH 5.0) to solubilize the 2S proteins; the precipitate of other proteins
- 25 removed by centrifugation. The acid extract was filtered using Microcon™ 0.2  $\mu\text{m}$  filtration units (Amicon Inc., Beverly, MA) and samples of the extracts from the transgenics were subjected to amino acid analysis and compared with the 2S albumin extracted from untransformed Arabidopsis (see Table 1).

- The seeds (T3) from those lines in the T2 generation that showed the most
- 30 increased Met and Lys content of the 2S fractions were sown to provide a T4 set of seeds for analysis. The seeds were treated as for the previous generation to obtain the 2S protein for analysis and the results shown in Table 1.

- Table 1 shows the percent by weight of Met and Lys in the 2S fraction of untransformed and transgenic Arabidopsis seeds of the T3 and T4 generations. The
- 35 percent by weight of Arg was included to indicate that concomitant decrease was observed with Lys increase, as expected. The percent by weight of valine (Val) is also

reported, a residue not present in the Brazil Nut 2S albumin, thus providing an internal reference for comparison of the various 2S extractions.

Table 1

The weight percent of Met and Lys in the 2S fraction of  
Arabidopsis transgenics producing the modified Brazil Nut 2S albumin

Arabidopsis variant	Seed Generation	%Met	%Lys	%Arg	%Val
C24	T3	2.2	7.9	9.8	2.7
BNwt	T3	3.6	7.3	10.3	2.6
BN15	T3	4.5	9.6	8.7	2.6
BN19	T3	4.5	10.4	8.2	2.8
BNwt	T4	6.7	7.5	11.6	3.3
BNwt	T4	6.5	7.4	11.7	3.3
BN15	T4	8.7	11.8	7.6	3.2
BN15	T4	10.0	13.1	7.5	3.2
BN19	T4	8.0	12.5	6.8	3.2
BN19	T4	8.0	12.5	7.2	3.1
BN19	T4	10.3	14.3	6.4	2.7
BN19	T4	10.4	14.4	6.3	2.7
BN19	T4	7.1	12.1	8.0	3.3
BN19	T4	7.0	11.9	8.0	3.3

5

### EXAMPLE 10

#### Expression of the Brazil Nut 2S Albumin Gene in *E. coli*

The vectors pETBNwt, pETBN15, pETBN16, pETBN17, pETBN18 and pETBN19 were used to transform *E. coli* (BL21) cells and grown in LB medium with kanamycin (30 ug/mL) selection in 50 mL shake cultures at 37° overnight on an incubated shaker (300 rpm). The next day, the cells were harvested to make glycerol stocks for long term storage and 1 mL was used to inoculate a fresh 50 mL batch of medium with the same selection. When the cells had reached an OD<sub>600</sub> of 0.9, protein expression was induced with 1 mM IPTG. The cells were harvested after overnight incubation on the shaker at 37°.

15

The protein content of the cells was analyzed by incubating a fraction of the cell paste that had been washed with 0.1M tris Cl buffer (pH 8.0) at 100° in a gel SDS

- loading buffer. Samples of the lysed cell extract were then run on a 18% SDS
- polyacrylamide gel. After the gel had been run it was allowed to wash in 0.1M CAPS buffer, pH 10, to remove the Tris-Glycine gel running buffer. The proteins were transferred to PVDF membranes using electrophoretic transblotting procedures and
- 5 visualized by coomassie blue staining. Those bands with the mobility of the 2S storage protein were identified as the recombinant product by N-terminal sequence and Western analyses.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
  - (B) STREET: 1007 MARKET STREET
  - (C) CITY: WILMINGTON
  - (D) STATE: DELAWARE
  - (E) COUNTRY: UNITED STATES OF AMERICA
  - (F) ZIP: 19898
  - (G) TELEPHONE: 302-992-5481
  - (H) TELEFAX: 302-773-0164
  - (I) TELEX: 6717325
- (ii) TITLE OF INVENTION: AN ENGINEERED HIGH SULFUR  
CONTAINING SEED PROTEIN CONTAINING  
OTHER ESSENTIAL AMINO ACIDS
- (iii) NUMBER OF SEQUENCES: 36
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: DISKETTE, 5.0 INCH
  - (B) COMPUTER: IBM PC COMPATIBLE
  - (C) OPERATING SYSTEM: MICROSOFT WINDOWS 95
  - (D) SOFTWARE: MICROSOFT WORD VERSION 7.0A
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 60/042,827
  - (B) FILING DATE: APRIL 8, 1997
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: CHRISTENBURY, LYNNE M.
  - (B) REGISTRATION NUMBER: 30,971
  - (C) REFERENCE/DOCKET NUMBER: BB-1069



## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 493 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 14..346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

GGGGAACCTT CAT ATG CAG GAG GAG TGT CGC GAG CAG ATG CAG AGA CAG      49
      Met Gln Glu Glu Cys Arg Glu Gln Met Gln Arg Gln
              1                      5                      10

CAG ATG CTC AGC CAC TGC CGG ATG TAC ATG AGA CAG CAG ATG GAG GAG      97
Gln Met Leu Ser His Cys Arg Met Tyr Met Arg Gln Gln Met Glu Glu
      15                      20                      25

AGC CCG TAC CAG ACC ATG CCC AGG CGG GGA ATG GAG CCG CAC ATG AGC      145
Ser Pro Tyr Gln Thr Met Pro Arg Arg Gly Met Glu Pro His Met Ser
      30                      35                      40

GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC AGA TGC GAA      193
Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys Arg Cys Glu
      45                      50                      55                      60

GGC TTA AGG ATG ATG ATG ATG AGG ATG CAA CAG GAG GAG ATG CAA CCC      241
Gly Leu Arg Met Met Met Met Arg Met Gln Gln Glu Glu Met Gln Pro
              65                      70                      75

CGA GGG GAG CAG ATG CGA AGG ATG ATG AGG CTG GCC GAG AAT ATC CCT      289
Arg Gly Glu Gln Met Arg Arg Met Met Arg Leu Ala Glu Asn Ile Pro
      80                      85                      90

TCC CGC TGC AAC CTC AGT CCC ATG AGA TGC CCC ATG GGT GGC TCC ATT      337
Ser Arg Cys Asn Leu Ser Pro Met Arg Cys Pro Met Gly Gly Ser Ile
      95                      100                      105

GCC GGG TTC TGAATCTGCC ACTAGCCAGT GCTGTAAATG TTAATAAGGC      386
Ala Gly Phe
      110

TCTCACAAAC TAGCTCTTTG TTGGCTTTTG GCCGGAGACT AGGGTGTGGG GAATTCGAGC      446

TCGGTACCCG GGGATCCTCT AGAGTCGACC TGCAGGCATG CAAGCTT      493

```

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 111 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Gln Glu Glu Cys Arg Glu Gln Met Gln Arg Gln Gln Met Leu Ser
 1             5             10             15

His Cys Arg Met Tyr Met Arg Gln Gln Met Glu Glu Ser Pro Tyr Gln
             20             25             30

Thr Met Pro Arg Arg Gly Met Glu Pro His Met Ser Glu Cys Cys Glu
             35             40             45

Gln Leu Glu Gly Met Asp Glu Ser Cys Arg Cys Glu Gly Leu Arg Met
 50             55             60

Met Met Met Arg Met Gln Gln Glu Glu Met Gln Pro Arg Gly Glu Gln
 65             70             75             80

Met Arg Arg Met Met Arg Leu Ala Glu Asn Ile Pro Ser Arg Cys Asn
             85             90             95

Leu Ser Pro Met Arg Cys Pro Met Gly Gly Ser Ile Ala Gly Phe
 100            105            110

```

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 28 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCAGACCATG CCGCGGAAGG GAATGGAG

28

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 34 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAGCTGCAAA TGCGAAGGCC TAAAGATGAT GATG

34

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGAGCAGAT GAAAAGATG ATGAAGCTAG CCGAGAATA

39

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTCCCATGAA ATGCCCTTC GGTGGATCGA TTGCCGGG

38

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 336 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..333

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATG CAG GAG GAG TGT CGC GAG CAG ATG CAG AGA CAG CAG ATG CTC AGC  
 Met Gln Glu Glu Cys Arg Glu Gln Met Gln Arg Gln Gln Met Leu Ser  
 1 5 10 15

48

CAC TGC CGG ATG TAC ATG AGA CAG CAG ATG GAG GAG AGC CCG TAC CAG	96
His Cys Arg Met Tyr Met Arg Gln Gln Met Glu Glu Ser Pro Tyr Gln	
20 25 30	
ACC ATG CCG CGG AAG GGA ATG GAG CCG CAC ATG AGC GAG TGC TGC GAG	144
Thr Met Pro Arg Lys Gly Met Glu Pro His Met Ser Glu Cys Cys Glu	
35 40 45	
CAG CTG GAG GGG ATG GAC GAG AGC TGC AAA TGC GAA GGC CTA AAG ATG	192
Gln Leu Glu Gly Met Asp Glu Ser Cys Lys Cys Glu Gly Leu Lys Met	
50 55 60	
ATG ATG ATG AGG ATG CAA CAG GAG GAG ATG CAA CCC CGA GGG GAG CAG	240
Met Met Met Arg Met Gln Gln Glu Glu Met Gln Pro Arg Gly Glu Gln	
65 70 75 80	
ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT TCC CGC TGC AAC	288
Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro Ser Arg Cys Asn	
85 90 95	
CTC AGT CCC ATG AAA TGC CCC TTC GGT GGA TCG ATT GCC GGG TTC	333
Leu Ser Pro Met Lys Cys Pro Phe Gly Gly Ser Ile Ala Gly Phe	
100 105 110	
TGA	336

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGGAAGCTT CATATGCAGG AGGAGTGTA AGAGCAGATG CAGAAACAGA AGATGCTCAG	60
CCACTGCAAG ATGTACATGA AACAGCAGAT GGAGGAGAGC CCGTACCAGA GCATGCCCGCG	120
GAAGGGAATG G	131

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

CCATTCCCTT CGCGGGCATG CTCTGGTACG GGCTCTCCTC CATCTGCTGT TTCATGTACA   60
TCTTGCACTG GCTGAGCATC TTCTGTTTCT GCATCTGCTC TTTACTACTCC TCCTGCATAT  120
GAAGCTTCCC C                                                              131

```

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 493 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 14..346

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

GGGGAACCTT CAT ATG CAG GAG GAG TGT AAA GAG CAG ATG CAG AAA CAG   49
      Met Gln Glu Glu Cys Lys Glu Gln Met Gln Lys Gln
        1              5              10

AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG ATG GAG GAG   97
Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln Met Glu Glu
    15              20              25

AGC CCG TAC CAG AGC ATG CCG CGG AAG GGA ATG GAG CCG CAC ATG AGC  145
Ser Pro Tyr Gln Ser Met Pro Arg Lys Gly Met Glu Pro His Met Ser
    30              35              40

GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC AAA TGC GAA  193
Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys Lys Cys Glu
    45              50              55              60

GGC CTA AAG ATG ATG ATG ATG AGG ATG CAA CAG GAG GAG ATG CAA CCC  241
Gly Leu Lys Met Met Met Met Arg Met Gln Gln Glu Glu Met Gln Pro
    65              70              75

CGA GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GGC GAG AAT ATC CCT  289
Arg Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro
    80              85              90

TCC CGC TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT GGA TCG ATT  337
Ser Arg Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly Gly Ser Ile
    95              100             105

GCC GGG TTC TGAATCTGCC ACTAGCCAGT GCTGTAAATG TTAATAAGGC      386
Ala Gly Phe
    110

```

TCTCACAAC TAGCTCTTTG TTGGCTTTTG GCCGGAGACT AGGGTGTGGG GAATTCGAGC 446  
TCGGTACCCG GGGATCTCTT AGAGTCGACC TGCAGGCATG CAAGCTT 493

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 151 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTACCAGAGC ATGCCGAAGA AGGGAATGGA GCCGCACATG AGCGAGTGCT GCGAGCAGCT 60  
GGAGGGGATG GACGAGAGCT GCAATGCGA AGGCCTAAAG ATGATGATGA TGAAGATGCA 120  
ACAGGAGGAG ATGCAACCCA AGGGGGAGCA G 151

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 141 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GATGCAACCC AAGGGGGAGC AGATGAAAAA GATGATGAAG CTAGCCGAGA ATATCCCTTC 60  
CAAAATGCAAC CTAGTCCCA TGAAATGCCC CTTCGGTGGA TCGATTGCCG GGTTCAGAG 120  
ATCCGAATTC AAGCTTGCGG C 141

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 141 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCCGCAAGCT TGAATTCGGA TCCTCAGAAC CCGGCAATCG ATCCACCGAA GGGGCATTTC 60

ATGGGACTGA GGTTCATTG GGAAGGGATA TTCTCGGCTA GCTTCATCAT CTTTTTCATC 120  
TGCTCCCCCT TGGGTTGCAT C 141

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 151 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEO ID NO:14:

CTGCTCCCC TTGGGTTGCA TCTCCTCTG TTGCATCTTC ATCATCATCA TCTTTAGGCC 60  
TTGCGATTTC CAGCTCTCGT CCATCCCCC CAGCTGCTCG CAGCACTGCG TCATGTGCGG 120  
CTCCATTCCC TTCTTCGGCA TGCTCTGGA C 151

(2) INFORMATION FOR SEO ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
(B) LOCATION: 14..346

(xi) SEQUENCE DESCRIPTION: SEO ID NO:15:

GGGGAACCTT	CAT	ATG	CAG	GAG	GAG	TGT	AAA	GAG	CAG	ATG	CAG	AAA	CAG		49	
	Met	Gln	Glu	Glu	Cys	Lys	Glu	Gln	Met	Gln	Lys	Gln				
	1				5					10						
AAG	ATG	CTC	AGC	CAC	TGC	AAG	ATG	TAC	ATG	AAA	CAG	CAG	ATG	GAG	GAG	97
Lys	Met	Leu	Ser	His	Cys	Lys	Met	Tyr	Met	Lys	Gln	Gln	Met	Glu	Glu	
	15				20						25					
AGC	CCG	TAC	CAG	AGC	ATG	CCG	AAG	AAG	GGA	ATG	GAG	CCG	CAC	ATG	AGC	145
Ser	Pro	Tyr	Gln	Ser	Met	Pro	Lys	Lys	Gly	Met	Glu	Pro	His	Met	Ser	
	30				35						40					
GAG	TGC	TGC	GAG	CAG	CTG	GAG	GGG	ATG	GAC	GAG	AGC	TGC	AAA	TGC	GAA	193
Glu	Cys	Cys	Glu	Gln	Leu	Glu	Gly	Met	Asp	Glu	Ser	Cys	Lys	Cys	Glu	
45					50					55					60	

GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG ATG CAA CCC	241
Gly Leu Lys Met Met Met Lys Met Gln Gln Glu Glu Met Gln Pro	
65 70 75	
AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT	289
Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro	
80 85 90	
TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT GGA TCG ATT	337
Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly Gly Ser Ile	
95 100 105	
GCC GGG TTC TGAGGATCCG AATCAAGCT T	367
Ala Gly Phe	
110	

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 100 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GATGATGAAG CTAGCCGAGA ATATCCCTTC CAAATGCAAC CTCAGTCCCA TGAAATGCCC	60
CTTCAAAGGA AAGATTGCCG GGTTCGAGG ATCCGAATTC	100

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 100 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GAATTCGGAT CCTCAGAACC CGGCAATCTT TCCTTTGAAG GGCATTTC A TGGGACTGAG	60
GTTGCATTTC GAAGGGATAT TCTCGGCTAG CTTCATCATC	100

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 367 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear



(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 14..346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

GGGGAACCTT CAT ATG CAG GAG GAG TGT AAA GAG CAG ATG CAG AAA CAG      49
      Met Gln Glu Glu Cys Lys Glu Gln Met Gln Lys Gln
      1                      5                      10

AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG ATG GAG GAG      97
Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln Met Glu Glu
      15                      20                      25

AGC CCG TAC CAG AGC ATG CCG AAG AAG GGA ATG GAG CCG CAC ATG AGC      145
Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Pro His Met Ser
      30                      35                      40

GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC AAA TGC GAA      193
Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys Lys Cys Glu
      45                      50                      55

GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG ATG CAA CCC      241
Gly Leu Lys Met Met Met Met Lys Met Gln Gln Glu Glu Met Gln Pro
      65                      70                      75

AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT      289
Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro
      80                      85                      90

TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC AAA GGA AAG ATT      337
Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Lys Gly Lys Ile
      95                      100                      105

GCC GGG TTC TGAGGATCCG AATTCAGCT T      367
Ala Gly Phe
      110

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 105 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

CCGTACCAGA GCATCCCGAA GAAGGGAATG GAGCCGCACA TGAAAGAGTG CTGCGAGCAG      60

```



AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT 289  
 Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro  
                     80                                    85                                    90

TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC AAA GGA AAG ATT 337  
 Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Lys Gly Lys Ile  
                     95                                    100                                    105

GCC GGG TTC TGAGGATCCG AATCAAGCT T 367  
 Ala Gly Phe  
                     110

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGAGATATAC ATATGCAGGA GAAGTGTAAG GAGCAGATGC AGAAACAGAA GATGCTCAGC 60  
 CACTGCAAGA TGTACATGAA ACAGCAGATG ATGAAGAGCC CGTACCAGAG CATGCCGAAG 120  
 AAGCC 125

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCCTTCTTCG GCATGCTCTG GTACGGGCTC TTCATCATCT GCTGTTTCAT GTACATCTTG 60  
 CAGTGGCTGA GCATCTTCTG TTTCTGCATC TGCTCTTTAC ACTTCTCCTG CATATGTATA 120  
 TCTCC 125

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 14..346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```

GGGGAACCTT CAT ATG CAG GAG AAG TGT AAA GAG CAG ATG CAG AAA CAG      49
      Met Gln Glu Lys Cys Lys Glu Gln Met Gln Lys Gln
              1                      5                      10

AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG ATG ATG AAG      97
Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln Met Met Lys
      15                      20                      25

AGC CCG TAC CAG AGC ATG CCG AAG AAG GGA ATG GAG CCG CAC ATG AGC     145
Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Pro His Met Ser
      30                      35                      40

GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC AAA TGC GAA     193
Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys Lys Cys Glu
      45                      50                      55

GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG ATG CAA CCC     241
Gly Leu Lys Met Met Met Met Lys Met Gln Gln Glu Glu Met Gln Pro
      65                      70                      75

AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT     289
Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro
      80                      85                      90

TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT GGA TCG ATT     337
Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly Gly Ser Ile
      95                      100                      105

GCC GGG TTC TGAGGATCCG AATTCAAGCT T                                367
Ala Gly Phe
      110

```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 124 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```

GCGTACGACA TATGCAGGAG AAGGTAAAG AGCAGATGCA GAAACAGAAG ATGCTTAAGC      60

```

ACTGCAAGAT GTACATGAAA CAGCAGATGG AGGAGAGCCC GTACCAGAGC ATGCCGAAGA 120  
AGGG 124

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 124 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCTTCTTCG GCATGCTCTG GTACGGGCTC TCCTCCATCT GCTGTTTCAT GTACATCTTG 60  
CAGTGCTTAA GCATCTTCTG TTTCTGCATC TGCTCTTTAC ACTTCTCCTG CATATGTCGT 120  
ACGC 124

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 108 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCGTACCAGA GCATGCCGAA GAAGGGAATG GAGAAGTGA TGAGCGAGTG CTGCGAGCAG 60  
CTGGAGGGGA TGGACGAGAG CTGTAAATGC GAAGGCCTAA AGATGATG 108

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 108 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CATCATCTTT AGGCCTTCGC ATTTACAGCT CTCGTCCATC CCCTCGAGCT GCTCGCAGCA 60  
CTCGCTCATC CACTTCTCCA TTCCCTTCTT CGGCATGCTC TGGTACGG 108

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 367 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 14..346

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```

GGGGAACCTT CAT ATG CAG GAG AAG TGT AAA GAG CAG ATG CAG AAA CAG      49
      Met Gln Glu Lys Cys Lys Glu Gln Met Gln Lys Gln
          1                      5                      10

AAG ATG CTT AAG CAC TGC AAG ATG TAC ATG AAA CAG CAG ATG GAG GAG      97
Lys Met Leu Lys His Cys Lys Met Tyr Met Lys Gln Gln Met Glu Glu
      15                      20                      25

AGC CCG TAC CAG AGC ATG CCG AAG AAG GGA ATG GAG AAG TGG ATG AGC      145
Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Lys Trp Met Ser
      30                      35                      40

GAG TGC TGC GAG CAG CTC GAG GGG ATG GAC GAG AGC TGC AAA TGC GAA      193
Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys Lys Cys Glu
      45                      50                      55                      60

GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG ATG CAA CCC      241
Gly Leu Lys Met Met Met Met Lys Met Gln Gln Glu Glu Met Gln Pro
          65                      70                      75

AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG AAT ATC CCT      289
Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu Asn Ile Pro
          80                      85                      90

TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT GGA TCG ATT      337
Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly Gly Ser Ile
          95                      100                      105

GCC GGG TTC TGAGGATCCG AATTCAAGCT T      367
Ala Gly Phe
      110

```

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 139 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
GCAGTCAGCA CCATGGCGAA CAAGCTCTTC CTCGTCTGTG CGGCACTTGC CCTCTGCTTC   60
CTCCTTACGA ACGCGTCAAT TTACCGGACG GTCGTGGAGT TCGAGGAGGA CGACGCGACG   120
AATCATATGC GTTACAGCG                                     139
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 139 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
CGCTGTAACG CATATGATTC GTCGCGTCGT CCTCTCGAA CTCACGACC GTCGCGTAAA   60
TTGACGCGTT CGTAAGGAGG AAGCAGACGG CAAGTGCCGC ACAGACGAGG AAGAGCTTGT   120
TCGCCATGGT GCTGACTGC                                     139
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..449

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
CC ATG GCG AAC AAG CTC TTC CTC GTC TGT GCG GCA CTT GCC CTC TGC   47
Met Ala Asn Lys Leu Phe Leu Val Cys Ala Ala Leu Ala Leu Cys
  1             5             10             15

TTC CTC CTT ACG AAC GCG TCA ATT TAC CGG ACG GTC GTG GAG TTC GAG   95
Phe Leu Leu Thr Asn Ala Ser Ile Tyr Arg Thr Val Val Glu Phe Glu
          20             25             30
```

GAG GAC GAC GCG ACG AAT CAT ATG CAG GAG GAG TGT AAA GAG CAG ATG	143
Glu Asp Asp Ala Thr Asn His Met Gln Glu Glu Cys Lys Glu Gln Met	
35 40 45	
CAG AAA CAG AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG	191
Gln Lys Gln Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln	
50 55 60	
ATG GAG GAG AGC CCG TAC CAG AGC ATG CCC AAG AAG GGA ATG GAG CCG	239
Met Glu Glu Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Pro	
65 70 75	
CAC ATG AGC GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC	287
His Met Ser Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys	
80 85 90 95	
AAA TGC GAA GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG	335
Lys Cys Glu Gly Leu Lys Met Met Met Lys Met Lys Met Gln Gln Glu Glu	
100 105 110	
ATG CAA CCC AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG	383
Met Gln Pro Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu	
115 120 125	
AAT ATC CCT TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT	431
Asn Ile Pro Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly	
130 135 140	
GGA TCG ATT GCC GGG TTC TGAGGATCCG AATTCAGCT T	470
Gly Ser Ile Ala Gly Phe	
145	

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 470 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
- (A) NAME/KEY: CDS
  - (B) LOCATION: 3..449

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CC ATG GCG AAC AAG CTC TTC CTC GTC TGT GCG GCA CTT GCC CTC TGC	47
Met Ala Asn Lys Leu Phe Leu Val Cys Ala Ala Leu Ala Leu Cys	
1 5 10 15	
TTC CTC CTT ACG AAC GCG TCA ATT TAC CGG ACG GTC GTG GAG TTC GAG	95
Phe Leu Leu Thr Asn Ala Ser Ile Tyr Arg Thr Val Val Glu Phe Glu	
20 25 30	



GAG GAC GAC GCG ACG AAT CAT ATG CAG GAG AAG TGT AAA GAG CAG ATG	143
Glu Asp Asp Ala Thr Asn His Met Gln Glu Lys Cys Lys Glu Gln Met	
35 40 45	
CAG AAA CAG AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG	191
Gln Lys Gln Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln	
50 55 60	
ATG ATG AAG AGC CCG TAC CAG AGC ATG CCC AAG AAG GGA ATG GAG CCG	239
Met Met Lys Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Pro	
65 70 75	
CAC ATG AGC GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC	287
His Met Ser Glu Cys Cys Glu Gln Leu Glu Gly Met Asp Glu Ser Cys	
80 85 90 95	
AAA TGC GAA GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG	335
Lys Cys Glu Gly Leu Lys Met Met Met Met Lys Met Gln Gln Glu Glu	
100 105 110	
ATG CAA CCC AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC GAG	383
Met Gln Pro Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Glu	
115 120 125	
AAT ATC CCT TCC AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TTC GGT	431
Asn Ile Pro Ser Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Phe Gly	
130 135 140	
GGA TCG ATT GCC GGG TTC TGAGGATCCG AATCAAGCT T	470
Gly Ser Ile Ala Gly Phe	
145	

## (2) INFORMATION FOR SEQ ID NO:34:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATGATGAAG CTAGCCTGGA ATATCCCTTG GAAATGCAAC CTCAGTCCCA TGAAATGCCC	60
CTGGGGTGGA AAGATTGCCG GGTTCTGACC GCGGATCCGA ATCAAGCTT ACGTAACGAC	120

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GTCTGTACGT AAGCTTGAAT TCGGATCCGC GGTCAGAACC CGGCAATCTT TCCACCCAG 60

GGGCATTTC TGGGACTGAG GTTGCAATTC CAAGGGATAT TCCAGGCTAG CTTTCATCATC 120

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 474 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 3..449

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CC ATG GCG AAC AAG CTC TTC CTC GTC TGT GCG GCA CTT GCC CTC TGC 47  
Met Ala Asn Lys Leu Phe Leu Val Cys Ala Ala Leu Ala Leu Cys  
1 5 10 15

TTC CTC CTT ACG AAC GCG TCA ATT TAC CGG ACG GTC GTG GAG TTC GAG 95  
Phe Leu Leu Thr Asn Ala Ser Ile Tyr Arg Thr Val Val Glu Phe Glu  
20 25 30

GAG GAC GAC GCG ACG AAT CAT ATG CAG GAG GAG TGT AAA GAG CAG ATG 143  
Glu Asp Asp Ala Thr Asn His Met Gln Glu Glu Cys Lys Glu Gln Met  
35 40 45

CAG AAA CAG AAG ATG CTC AGC CAC TGC AAG ATG TAC ATG AAA CAG CAG 191  
Gln Lys Gln Lys Met Leu Ser His Cys Lys Met Tyr Met Lys Gln Gln  
50 55 60

ATG GAG GAG AGC CCG TAC CAG AGC ATG CCC AAG AAG GGA ATG GAG CCG 239  
Met Glu Glu Ser Pro Tyr Gln Ser Met Pro Lys Lys Gly Met Glu Pro  
65 70 75

CAC ATG AGC GAG TGC TGC GAG CAG CTG GAG GGG ATG GAC GAG AGC TGC 287  
His Met Ser Glu Cys Glu Glu Gln Leu Glu Gly Met Asp Glu Ser Cys  
80 85 90 95

AAA TGC GAA GGC CTA AAG ATG ATG ATG ATG AAG ATG CAA CAG GAG GAG 335  
Lys Cys Glu Gly Leu Lys Met Met Met Met Lys Met Gln Gln Glu Glu  
100 105 110

ATG CAA CCC AAG GGG GAG CAG ATG AAA AAG ATG ATG AAG CTA GCC TGG	383
Met Gln Pro Lys Gly Glu Gln Met Lys Lys Met Met Lys Leu Ala Trp	
115 120 125	
AAT ATC CCT TGG AAA TGC AAC CTC AGT CCC ATG AAA TGC CCC TGG GGT	431
Asn Ile Pro Trp Lys Cys Asn Leu Ser Pro Met Lys Cys Pro Trp Gly	
130 135 140	
GGA AAG ATT GCC GGG TTC TGACCGCGGA TCCGAATTCA AGCTT	474
Gly Lys Ile Ala Gly Phe	
145	

CLAIMS

- What is claimed is:

1. A modified Brazil Nut 2S albumin seed storage protein wherein:
  - (i) the amino acid sequence of the modified protein is at least 40% homologous to the wild type Brazil Nut 2S albumin seed storage protein;
  - (ii) all cysteine residues of the modified protein are conserved relative to the wild type protein;
  - (iii) at least 40% of proline residues are conserved relative to the wild type protein;
  - 10 (iv) at least 80% of leucine residues are conserved relative to the wild type protein; and
  - (v) the modified protein comprises at least one non-conservative amino acid substitution not within the hypervariable loop, the substitution consisting of replacement of a non-essential amino acid with an essential amino acid.
- 15 2. The modified Brazil Nut 2S albumin seed storage protein of Claim 1 wherein:
  - (i) the amino acid sequence of the modified protein is at least 82% homologous to the wild type Brazil Nut 2S albumin seed storage protein;
  - (ii) all cysteine, proline, leucine and methionine residues of the modified  
20 protein are conserved relative to the wild type protein;
  - (iii) all arginine residues of the wild type protein are substituted with lysine residues;
  - (iv) the modified protein comprises at least three non-conservative amino acid substitutions not within the hypervariable loop, said substitutions comprising  
25 substituting two glutamic acid residues with lysine residues and substituting one glutamine residue with a lysine residue.
3. The modified Brazil Nut 2S albumin seed storage protein of Claim 1 wherein the protein is a member of the group consisting of BNCNSS, BN11, BN15, BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN19 and AT2S1BN153W.
- 30 4. An isolated nucleic acid fragment encoding the modified Brazil Nut 2S albumin seed storage protein of Claim 1.
5. The nucleic acid fragment of Claim 4 comprising a nucleotide sequence identical or substantially similar to a member selected from the group consisting of SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:21, SEQ ID  
35 NO:24, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:33 and SEQ ID NO:36.
6. The nucleic acid fragment of Claim 5 comprising a nucleotide sequence selected from the group consisting nucleotide sequences that encode the amino acid

sequence set forth in SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:24, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:36.

7. A chimeric gene comprising the nucleic acid fragment of Claim 4, Claim 5 or Claim 6, operably linked to suitable regulatory sequences.

8. A transformed plant comprising in its genome the chimeric gene of Claim 7.

9. The transformed plant of Claim 8 wherein the plant is a member of the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, rice and forage grasses.

10. 10. Seeds derived from the transformed plant of Claim 8 wherein the seeds comprise the chimeric gene.

11. A method for increasing the essential amino acid content of seeds comprising:

(a) preparing a nucleic acid fragment encoding a modified Brazil Nut 2S albumin seed storage protein wherein:

(i) the amino acid sequence of the modified protein is at least 40% homologous to the wild type Brazil Nut 2S albumin seed storage protein;

(ii) all cysteine residues of the modified protein are conserved relative to the wild type protein;

20 (iii) at least 40% of proline residues are conserved relative to the wild type protein;

(iv) at least 80% of leucine residues are conserved relative to the wild type protein; and

25 (v) the modified protein comprises at least one non-conservative amino acid substitution not within the hypervariable loop, the substitution consisting of replacement of a non-essential amino acid with an essential amino acid;

(b) preparing a chimeric gene comprising the nucleic acid fragment of step(a) operably linked to suitable regulatory sequences;

(c) transforming a plant with the chimeric gene of step (b); and

30 (d) obtaining seeds from the transformed plant of step (c).

1/23

## FIG. 1

NdeI  
 GGGGAACCTTCATATGCGAGGAGGTGTCGCGAGCAGATGCGAGAGCAGACGATGCTCAG  
 1 +-----+-----+-----+-----+-----+  
 CCCCTGGAAAGTATACGTCCTCCTCACAGCGTGTCTACGTCCTGTCTGTACGAGTC  
 +-----+-----+-----+-----+-----+ 60  
  
 M Q E E C R E Q M Q R Q Q M L S  
 CCACTGCCGATGTACATGACAGACAGCATGGAGGAGCCGTACACAGACCATGCCAG  
 61 +-----+-----+-----+-----+-----+ 120  
 GGTACGGCCTACATGTACTCTGTCTACCTCCTCTCGGCATGGTCTGGTACGGGTC  
  
 H C R M Y M R Q Q M E E S P Y Q T M P R  
 GCGGGAAATGGAGCCGCACATGAGCGAGTGCTGCGAGCAGCTGGAGGGGATGGACAGAG  
 121 +-----+-----+-----+-----+-----+ 180  
 CGCCCCCTACCTCGGCGTGTACTCGCTCACGACGCTCGTCGACCTCCCCCTACCTGCTCTC  
  
 R G M E P H M S E C E Q L E G M D E S  
 PstI  
 CTCGAAATGCGAAGGTTAAGGATGATGATGATGAGGATGCAACAGGAGGAGATGCAACC  
 181 +-----+-----+-----+-----+-----+ 240  
 GACGCTACGCTTCGGAATTCCTACTACTACTACTCTCCTACGTTGCTCCTCTCTACGTTGG  
  
 C R C E G L R M M M M R M Q Q E E M Q P  
 CCGAGGGAGCAGATGCGAAGGATGATGAGCTGCCCGAGAATATCCCTTCCCCTGCAA  
 241 +-----+-----+-----+-----+-----+ 300  
 GGCTCCCCTCGTCTACGCTTCCTACTACTCGGACGGGCTCTTATAGGAGGGGCGACGTT  
  
 R G E Q M R R M M R L A E N I P S R C N

2/23

## FIG. 1 (Continued)

```

                                NcoI
CCTCAGTCCCATGAGATGCCCATGGGTGGCTCCATTGCCGGGTTCTGAATCTGCCACTA
301 -----+-----+-----+-----+-----+-----+ 360
GGAGTCAGGGTACTCTACGGGGTACCCACCGAGGTAACGCCCAAGACTTAGACGGTGAT

L S P M R C P M G G S I A G F *

GCCAGTGTGTAATGTTAATAAGGCTCTCACAACTAGCTCTTTGTTGCTTTTGCCG
361 -----+-----+-----+-----+-----+-----+ 420
CGGTCACGACATTTACAATTATTCGAGAGTGTTGATCGAGAAACAACCGAAACCGGC

                                EcoRI
GAGACTAGGTTGTGGGAATTCGAGCTCGGTACCCGGGATCCTCTAGAGTCGACCTGCA
421 -----+-----+-----+-----+-----+-----+ 480
CTCTGATCCACACCCCTTAAGCTCGAGCCATGGGCCCTAGGAGATCTCAGCTGGACGT

SphI HindIII
GGCATGCAAGCTT
481 -----+-----+-----+-----+-----+ 493
CCGTACGTTTCGAA

```

3/23

## FIG. 2

```

      ATGCAGGAGGAGTGTCCGAGCAGATGCAGAGACAGCAGATGCTACGCCACTGCCCGGATG
1  -----+-----+-----+-----+-----+-----+
   TAGTCTCTCTCACAGCGCTGCTACGTCTCTGTCTGTCTACGAGTCGGTGACGGCCCTAC
      M Q E E C R E Q M Q R Q Q M L S H C R M
                                     SacII
      TACATGAGACAGCAGATGGAGGAGAGCCCGTACCAGACCATGCCGCGGAAGGGAATGGAG
61 -----+-----+-----+-----+-----+-----+
   ATGTACTCTGTCTCTACCTCCTCTCGGGCATGGTCTGGTACGGCGCCCTGCCCTTACCTC
      Y M R Q Q M E E S P Y Q T M P R K G M E
                                     StuI
      CCGCACATGAGCGAGTGTCCGAGCAGCTGGAGGGGATGGACGAGAGCTGCAAAATGCGAA
121 -----+-----+-----+-----+-----+-----+
   GGCGTGCTACTCGCTCAGGACGCTCGTCGACCTCCCTACCTGCTCTCGACGTTTACGCTT
      P H M S E C C E Q L E G M D E S C K C E
                                     StuI
      GGCTTAAGATGATGATGAGGATGCAACAGGAGGAGATCAACCCCGAGGGGAGCAG
181 -----+-----+-----+-----+-----+-----+
   CCGGATTCTACTACTACTACTCTACGTTGTCTCTCTACGTTGGGGCTCCCTCGTC
      G L K M M M R M Q Q E E M Q P R G E Q

```



4/23

## FIG.2 (Continued)

NheI

241 ATGAAAGATGATGAAGCTAGCGAGAAATATCCCTCCCGCTGCAACCTCAGTCCCATG + 300  
 -----+-----+-----+-----+-----+-----+  
 TACTTTTCTACTACTTCGATCGGCTCTTATAGGAAGGCGGACGTTGGAGTCAGGTAC  
 M K K M M K L A E N I P S R C N L S P M

ClaI

301 AAATGCCCCCTTCGGTGGATCGGATGCCGGGTTCTGA + 336  
 -----+-----+-----+-----+-----+-----+  
 TTTACGGGGAAGCCACCTAGCTAACGGCCCCAAGACT  
 K C P F G G S I A G F \*

5/23

## FIG. 3

NdeI  
 1 GGGGAACCTTCATATGSCAGGAGGAGTGTAAAGAGCAGATGCAGAAACAGAGATGCTCAG 60  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 CCCCTGGAAGATACGTCTCCTCCTCACATTCTCGTCTACGTCTTTGCTTCTCTACGAGTC  
 M Q E E C K E Q M Q K Q K M L S

SphI SacII  
 61 CCACTGCAAGATGTACATGAACAGCAGATGGAGGAGAGCCCGTACCAGAGCATGCCGCG 120  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 GGTGACGTTCTACATGTACTTTGTGCTCTACCTCCTCTCGGGCATGGTCTCGTACGGCGC  
 H C K M Y M K Q Q M E E S P Y Q S M P R

StuI  
 121 GAAGGGAATGGAGCCGCACATGACGCGAGTGCTGCGAGCAGCTGGAGGGGATGGACGAGAG 180  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 CTTCCCTTACCTCGGCGTGACTCGCTCAGACGGCTCGTCGACCTCCCCCTACCTGCTCTC  
 K G M E P H M S E C C E Q L E G M D E S

181 CTGCAATGCGAAGGCTTAAAGATGATGATGAGGATGCAACAGGAGGAGATGCAACC 240  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 GACGTTACGCTTCGGATTCTACTACTACTCTCTACGTTGTCTCCTCTACGTTGG  
 C K C E G L K M M M M R M Q Q E E M Q P

6/23

## FIG. 3 (Continued)

```

                NheI
CCGGGGGAGCAGATGAAAAAGATGATGAAGCTAGCCGAGAAATATCCCTTCCCGCTGCAA
241 +-----+-----+-----+-----+-----+-----+ 300
GGTCCCTCGTCTACTTTTCTACTACTTCGATCGGCTCTTATAGGGAAGGGCGACGTT
R G E Q M K K M M K L A E N I P S R C N

                ClaI
CCTCAGTCCCATGAATGCCCTTCGGGTGGATCGATTGCCGGGTCTCTGAATCTGCCACTA
301 +-----+-----+-----+-----+-----+-----+ 360
GGAGTCAGGTACTTTACGGGGAGCCACCTAGCTACGGCCCAAGACTTAGACGGTGAT
L S P M K C P F G G S I A G F *

GCCAGTGTGTAATGTTAATAAGGCTCTCACAACTAGCTCTTTGTTGGCTTTTGGCCG
361 +-----+-----+-----+-----+-----+-----+ 420
CGGTCACGACATTTACAATTATTCCGAGAGTGTTGATCGAGAAACAACCGAAAACCGGC

                EcoRI
GAGACTAGGTGTGGGAATTCGAGCTCGGTACCCGGGGATCCCTCTAGAGTCGACCTGCA
421 +-----+-----+-----+-----+-----+-----+ 480
CTCTGATCCACACCCCTTAAGCTCGAGCCATGGGCCCCCTAGGAGATCTCAGCTGGACGT

                SphI HindIII
GGCATGCAAGCTT
481 +-----+-----+-----+-----+-----+-----+ 493
CCGTACGTTTCGAA

```

7/23

FIG. 4

NdeI  
 1 GGGGAACCTTCATATGAGGAGGAGTGTAAAGAGCAGATGCAGAAACAGAGATGCTCAG  
 -----+-----+-----+-----+-----+  
 60 CCCCTTGAAGTATACGTCCTCCTCACAATTCTCGTCTACGCTTTGTCTTCTACGAGTC  
 -----+-----+-----+-----+-----+  
 M Q E E C K E Q M Q K Q K M L S

SphI  
 61 CCACTGCAAGATGTACATGAACAGCAGATGGAGGAGGCCGTACCCAGAGCATGCCGAA  
 -----+-----+-----+-----+-----+  
 120 GGTGACGTTCTACATGTACTTGTGCTCTACCTCCTCTCGGGCATGGTCTCGTACGGGCT  
 -----+-----+-----+-----+-----+  
 H C K M Y M K Q Q M E E S P Y Q S M P K

121 GAAGGAATGGAGCCGCACATGAGCGAGTCTGCGAGCAGCTGGAGGGGATGGACGAGAG  
 -----+-----+-----+-----+-----+  
 180 CTCCCTTACCTCGGCGTGACTCGCTCAGGACGCTCGTCGACCTCCCGTACCTGCTCTC  
 -----+-----+-----+-----+-----+  
 K G M E P H M S E C C E Q L E G M D E S

StuI  
 181 CTGCAATGCGAAGGCCTTAAGATGATGATGAAGATGCAACAGGAGGAGATGCAACC  
 -----+-----+-----+-----+-----+  
 240 GACGTTACGCTTCGGGATTCTACTACTACTACTTCTACGTTGTCTCCTCTACGTTGG  
 -----+-----+-----+-----+-----+  
 C K C E G L K M M M M K M Q Q E E M Q P

8/23

FIG. 4 (Continued)

StyI                  NheI

CAAGGGGGAACAGATGAAAAAAGATGATGAAGCTAGCGCAGAATAATCCCTTCCAATGCAA

-----+-----+

241

GTTCCCCCTCGTGACTACTTTTCGTACTACTTCGATCGGCTCTTATAGGAAGGTTTTACGTT

-----+-----+

300

K G E Q M K K M M K L A E N I P S K C N

ClaI BamHI EcoRI  
CCTCAGTCCCATGAATGCCCTTCGTGGATCGATTGCGGGTTCTGAGGATCCGAATT  
+ + + + +  
GGAGTCAGGGTACTTACGGGGAAACCACCTAGCTAACGGCCCCAAGCACTCCTAGGCCTAA  
301 + + + + + 360

L S P M K C P F G S I A G E \*

HindIII  
CAAGCTT  
-----  
GTTCGAA  
361

9/23

## FIG. 5

NdeI  
 GGGGRACCTTCATATGCAGGAGGAGTGTAAAGAGCAGATGCAGAAACAGAAAGATGCTCAG  
 1 -----+-----+-----+-----+-----+  
 CCCCTTGGGAAGTATACGTCCTCCTCACATTCTCGTACGTCCTTTGCTCTTCTACGAGTC  
 60  
 M Q E E C K E Q M Q K Q K M L S

SphI  
 CCACTGCAAGATGTACATGAACACAGCAGATGGAGGAGAGCCCGTACCAGAGCATGCCGAA  
 61 -----+-----+-----+-----+-----+  
 GGTGACGTTCTACATGTACTTTGTGCTCTACCTCCTCGGGCATGGTCTCGTACGGCTT  
 120  
 H C K M Y M K Q Q M E E S P Y Q S M P K

GAAGGGAATGGAGCCGACATGAGCGAGTGCTGCGAGCAGCTGGAGGGGATGGACGAGAG  
 121 -----+-----+-----+-----+-----+  
 CTTCCCTTACCTCGGGCGTGTACTCGCTCAGCAGCTCGTCGACCTCCCTACCTGCTCTC  
 180  
 K G M E P H M S E C C E Q L E G M D E S

StuI  
 CTGCAATGCGAAGGCCTAAAGATGATGATGAAGATGCAACAGGAGGAGATGCAACC  
 181 -----+-----+-----+-----+-----+  
 GACGTTACGCTTCGGATTCTCTACTACTACTTCTACGTTGTCTCCTCTACGTTTG  
 240  
 C K C E G L K M M M M K M Q Q E E M Q P

10/23

FIG. 5 (Continued)

StyI                      NheI  
CAAGGGGAGCAGATGA<sup>+</sup>AAAAAGATGATGAAGCTAGCGAGAATATCCCTTCCAATGCAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
241 GTCCCCCTCGTC<sup>-</sup>TACTTTTCTACTACTTCGATCGSCTCTTATAGGGAAGGTTTACGTT  
K G G E Q M K K M K L A E N I P S K C N  
  
BamHI EcoRI  
CCTCAGTCCCATGAATGCCCTTCAAAGGAAGATTCCGGGTCTGAGGATCCGAATT  
-----+-----+-----+-----+-----+-----+-----+-----+  
301 GGAGTCAGGTACTTTACGGGGAAGTTCTCTTAACGGCCCCAGACCTCCTAGGCTTAA  
L S P M K C P F K G K I A G F \*

11/23

FIG. 6

NdeI  
 1 GGGGAACCTTTCATATGCAGGAGGAGTGTAAAGAGCAGATGCAGAAACAGAAAGATGCTCAG  
 -----+-----+-----+-----+-----+-----+  
 60 CCCCTTGGAAAGTATACGTCCTCCTCACATTTTCTCGTCTACGTCCTTTGTCTTCTACGAGTC  
 -----+-----+-----+-----+-----+-----+

M Q E E C K E Q M Q K Q K M L S  
 -----+-----+-----+-----+-----+

SphI  
 61 CCACTGCAAGATGTACATGAACAGCAGATGGAGAGAGCCCGTACCAGAGCATGCCGAA  
 -----+-----+-----+-----+-----+-----+  
 120 GGTGACGTTCTACATGACTTTGTGCTACCTCCTCTCGGGCATGGTCTCGTACGGCTT  
 -----+-----+-----+-----+-----+-----+

H C K M Y M K Q Q M E S P Y Q S M P K  
 -----+-----+-----+-----+-----+-----+

121 GAAGGAATGGAGCGCACATGAAGAGTCTCGAGCAGCTGGAGGGGATGGACGAGAG  
 -----+-----+-----+-----+-----+-----+  
 180 CTTCCCTTACCTCGGCGTGACTTTCTCAGAGCTCGTGACCTCCCTACCTGCTCTC  
 -----+-----+-----+-----+-----+-----+

K G M E P H M K E C E Q L E G M D E S  
 -----+-----+-----+-----+-----+-----+

StuI  
 181 CTGCAATGCCGAGGCCCTAAGATGATGATGAAGATGAACAGAGGAGATGCAACC  
 -----+-----+-----+-----+-----+-----+  
 240 GACGTTTACGCTCCGGATTTCTACTACTACTTCTACGTTGTCTCCTCTACCTTGG  
 -----+-----+-----+-----+-----+-----+

C K C E G L K M M M K M Q Q E E M Q P  
 -----+-----+-----+-----+-----+-----+



12/23

## FIG. 6 (Continued)

StyI NheI  
CAAGGGGACAGATGAAGAAGATGATGAAGCTAGCCGAGATATCCTTCCAATGCAA  
-----+-----+-----+-----+-----+-----+  
241 GTTCCCCTCGTCTACTTTTCTACTACTTCGATCGGCTCTTATAGGGAAGGTTTAGGT  
  
K G E Q M K K M K L A E N I P S K C N

BamHI EcoRI  
CCTCAGTCCCATGAATGCCCTTCAAAGGAAGATTCCGGGTTC TGAGGATCCGAATT  
-----+-----+-----+-----+-----+-----+  
301 GGAGTCAGGTACTTTACGGGGAACTTTCCTTCTTAAGSCCCCAAGACTCCTAGSCTTAA  
  
L S P M K C P F K G K I A G F \*

HindIII  
CAAGCTT  
-----  
GTTCGAA  
361

FIG. 7

NdeI  
 GGGGAACCTTCATATGCAGGAGAAAGTGTAAAGAGCAGATGCAGAAACAGAAAGATGCTCAG 60  
 -----+-----+-----+-----+-----+  
 CCCCITGGAGTATACGTCTCTTCACATTTCTCGTCTACGTCTTTGCTCTCTACGAGTC  
 M Q E K C K E Q M Q K Q K M L S  
  
 SphI  
 CCACTGCAAGATGTACATGAACACAGCAGATGATGAAGAGCCCGTACCAGAGCATGCCGAA 120  
 -----+-----+-----+-----+-----+  
 GGTGACGTTCTACATGTACTTTGTCTCTACTACTTCTCGGCATGGTCTCTGACGGCTT  
 H C K M Y M K Q Q M M K S P Y Q S M P K  
  
 GAAGGGAATGGAGCGGCACATGAGCGAGTGTCTCGACGAGCTGGAGGGGATGGACGAGAG 180  
 -----+-----+-----+-----+-----+  
 CTTCCCTTACCTCGGCGTGTACTCGCTCAGGAGCTCGTCGACCTCCCTACCTACCTGCTCTC  
 K G M E P H M S E C C E Q L E G M D E S  
  
 StuI  
 CTGCAATCGGAGGCCATAAGATGATGATGAAGATGCAACAGGAGGAGATGCAACC 240  
 -----+-----+-----+-----+-----+  
 GACGTTACGCTTCGGATTCTACTACTACTACTACTCTACGTTCCTCCTCCTACGTTGG  
 C K C E G L K M M M M K M Q Q E E M Q P

13/23









18/23

FIG. 10

	NcoI	EarI	
	GCATGGCGAACAAAGTCTTCTCGTCTGTGCGGCACTTGCCTCTGCTTCTCTCTTACGA		60
1	GGTACCGCTTGTTCGAGAAGGAGCAGACACGCCGCTGAACGGGAGACGAAGGAGGAATGCT		
	M A N K L F L V C A A L A L C F L L T N		
	MluI	NdeI	
	ACGGTCAATTACCGGACGGTCGTGGAGTTCGAGGAGGACGACGCGACGAATCATATGC		120
61	TGCGCAGTTAAATGGCCTGCCAGCACCTCAAGCTCCTCTGCTGCGCTGCTTAGTATACG		
	A S I Y R T V V E F E E D D A T N H M Q		
	AGGAGGAGTGTAAAGAGCAGATGCAGAAACAGAAAGATGCTCAGCCACTGCAAGATGTACA		180
121	TCTCTTCACATTCTCGTCTACGTCTTTGTCTTACGAGTCGGTGACGTTCTTACATGT		
	E E C K E Q M Q K Q K M L S H C K M Y M		
	SphI		
	TGAAACAGCAGATGGAGGAGCCCGTACCAGAGCATGCCAAGAAGGAATGGAGCCGC		240
181	ACTTTGCTGTACTACTTCTCGGCGATGGTCTCGTAGCGGTTCTTCCCTTACCTCGGCG		
	K Q Q M E E S P Y Q S M P K K G M E P H		





20/23

FIG. 11

NcoI  
 CCATGGCGAACAAGCTCTTCTCGTCTGTGCGGCACTTGCCTCTGCTTCTCTCTACGA  
 1  
 GGTACCGCTTGTTCGAGAGGAGCAGACACGCCGTGAACGGGAGACGAAGGAGGAATGCT  
 60  
 M A N K L F L V C A A L A L C F L L T N  
 MluI  
 AGCGGTCAATTTACCGGACGGTCTGGAGTTCGAGGAGGACGACGCGACGAATCATATGC  
 61  
 TCGGCAGTTAAATGGCTGCCAGCACCTCAAGCTCCTCCTGCTGCGCTGCTTAGTATACG  
 120  
 A S I Y R T V V E F E D D A T N H M Q  
 AGGAGAGTGTAAAGAGCAGATGCAGAAACAGAAAGATGCTCAGCCACTGCAAGATGTACA  
 121  
 TCCTCTTCACATTTCTCGTCTACGCTTTGTCTCTACGAGTCGGTGACGTTCTACATGT  
 180  
 E K C K E Q M Q K K Q K M L S H C K M Y M  
 SphI  
 TGAACACGAGATGATGAAGAGCCCGTACCAGAGCATGCCCAAGAGGGAATGGAGCCGC  
 181  
 ACTTTGCTCTACTACTTCTCGGGCATGGTCTCGTACGGGTCTTCCCTTACCTCGGCG  
 240  
 K Q Q M M K S P Y Q S M P K K G M E P H

21/23

## FIG. 11(Continued)

241 ACATGAGCGAGTGCTCGAGCAGCTGGAGGGGATGGACGAGAGCTGCAAAATCGGAAGGCC <sup>StuI</sup>  
 -----+-----+-----+-----+-----+-----+  
 300 TGTACTCGCTACGACGCTCGTGCACCTCCCTACCTGCTCTCGACGTTTACGCTTCGGA  
 M S E C C E Q L E G M D E S C K C E G L

301 TAAAGATGATGATGAAGATGCAACAGGAGGAGATGCAACCCAGGGGAGCAGATGA <sup>StuI</sup>  
 -----+-----+-----+-----+-----+-----+  
 360 ATTTCTACTACTACTTCTAGTTGCTCCTCTACGTTGGGTTCCCTCGTCTACT  
 K M M M M K M Q Q E E M Q P K G E Q M K

361 AAAAGATGATGAAGCTAGCCGAGAAATATCCCTTCCAAATGCAACCTCAGTCCCATGAAAT <sup>NheI</sup>  
 -----+-----+-----+-----+-----+-----+  
 420 TTTTCTACTACTTCGATCGGCTCTATAGGAAGGTTTACGTTGGAGTCAGGGTACTTTA  
 K M M K L A E N I P S K C N L S P M K C

421 GCCCCTCGGTGGATCGATGCCGGGTTCTGAGGATCGGAATTCAGCTT <sup>StuI</sup> BamHI EcoRI HindIII  
 -----+-----+-----+-----+-----+-----+  
 CGGGGAAGCCACCTAGCTAAGGGCCCAAGACTCCTAGGCTTAAGTTCGAA

P F G G S I A G F \*

22/23

FIG. 12

```

      NcoI           EarI
      CCATGGCGAACAAAGCTCTTCCCTCGTCTGTGGGCACTTGGCCCTCTGCTTCCTCCTTACGA
1  -----+-----+-----+-----+-----+-----+-----+-----+
      GGTACCGCTTGTCGAGAAGGAGCAGACAGCCGTGAACGGGAGACGAGGAGGAATGCT      60
      M A N K L F L V C A A L A L C F L L T N
      MluI
      ACGCGTCAATTTACGGACGGTCGTGGAGTTCGAGGAGGACGACGCGACGAATCATATGC      NdeI
121 -----+-----+-----+-----+-----+-----+-----+-----+
      TGCGCAGTTAAATGGCCTGCCAGCACCTCAAGCTCCTCCTGCTGCGCTGCTTAGTATACG
      A S I Y R T V V E F E E D D A T N H M Q
      AGGAGGAGTGCTAAAGACGACAGATGCAGAAACAGAAGATGCTCAGCCACTGCAAGATGTACA      180
181 -----+-----+-----+-----+-----+-----+-----+-----+
      TCCTCTTCACATTTCTCGTCTACGCTTTGTCTTCTACGAGTCGGTGACGTTCTTACATGT
      E E C K E Q M Q K K Q K M L S H C K M Y M
      SphI
      TGAACACAGCAGATGGAGGAGAGCCCGTACCAGAGCATGCCCAAGRAGGGNATGGAGCCGC
181 -----+-----+-----+-----+-----+-----+-----+-----+
      ACTTTGTCGTCTACTTCTCGGGCATGGTCTCGTACGGGTTCTTCCCTTACCTCGGCG
      K Q Q M E E S P Y Q S M P K K G M E P H

```

23/23

## FIG. 12(Continued)

```

                StuI
ACATGAGCGAGTCTGCGAGCAGCTGGAGGGGATGACGAGAGCTGCAATGCGAAGGCC
241  -----+-----+-----+-----+-----+
TGTACTCGCTCACGACGCTCGTCGACTCCCTACTCTCTCGAGCTTACGCTTCCGA
                M S E C C E Q L E G M D E S C K C E G L

                StuI
TAAAGATGATGATGAAGATGCAACAGGAGGAGATGCAACCCAGGGGGAGCAGATGA
301  -----+-----+-----+-----+-----+
ATTTCTACTACTACTTCTACGTTGTCTCTCTCTACGTTGGGTTCCCTCGTCTACT
                K M M M M K M Q Q E E M Q P K G E Q M K

                NheI
AAAAGATGATGAAGCTAGCCTGGAAATATCCCTTGGAAATGCAACCTCAGTCCCATGAAAT
361  -----+-----+-----+-----+-----+
TTTTCTACTACTTCGATCGGACCTTATAGGGAACCTTTACGTTGGAGTCAGGCTACTTTA
                K M M K L A W N I P W K C N L S P M K C

                StuI
GCCCTGGGCTGGAAGATGCGGGTTCGTGACCGCGGATCCGAATTCAGCTT
421  -----+-----+-----+-----+-----+
CGGGGACCCACCTAGCTAACGCCCAAGACTCCTAGGCTTAAGTTCGAACGAA
                P W G G K I A G F *

```

## INTERNATIONAL SEARCH REPORT

Int. Application No.

PCT/US 98/06673

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/82 C12N15/29 A01H5/00

According to international Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARCELLINO L H ET AL: "Modified 2S albumins with improved tryptophan content are correctly expressed in transgenic tobacco plants." FEBS LETTERS 385 (3). 1996. 154-158. ISSN: 0014-5793, XP002071371 see the whole document ---	1,4,7-11
A	EP 0 318 341 A (PLANT GENETIC SYSTEMS NV) 31 May 1989 cited in the application see the whole document ---	1-11
-/--		

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document published prior to the international filing date but by other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

14 July 1998

Date of mailing of the international search report

28/07/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Maddox, A

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/06673

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No.
A	DE CLERCQ, A., ET AL.: "Stable accumulation of modified 2S albumin seed storage proteins with higher methionine contents in transgenic plants" PLANT PHYSIOLOGY, vol. 94, 1990, pages 970-979, XP002071372 cited in the application see the whole document ---	1-11
A	WO 94 10315 A (PIONEER HI BRED INT) 11 May 1994 cited in the application see the whole document ---	1-11
A	WO 94 16078 A (PIONEER HI BRED INT) 21 July 1994 see the whole document ---	1-11
A	RICE J A ET AL: "EXPRESSION OF SYNTHETIC HIGH LYSINE SEED STORAGE PROTEINS CAN SIGNIFICANTLY INCREASE THE ACCUMULATED LEVELS OF LYSINE IN MATURE SEEDS OF TRANSGENIC CROP PLANTS" JOURNAL OF CELLULAR BIOCHEMISTRY, vol. 18A, 9 January 1994, page 107 XP002035501 see abstract ---	1-11
A	WO 95 15392 A (DU PONT ;FALCO SAVERIO CARL (US); KEELER SHARON JO (US); RICE JANE) 8 June 1995 see the whole document ---	1-11
A	EP 0 319 353 A (PLANT GENETIC SYSTEMS NV) 7 June 1989 see the whole document ---	1-11
A	RADKE S E ET AL: "TRANSFORMATION OF BRASSICA NAPUS L. USING AGROBACTERIUM TUMEFACIENS: DEVELOPMENTALLY REGULATED EXPRESSION OF A REINTRODUCED NAPIN GENE" THEORETICAL AND APPLIED GENETICS, vol. 75, no. 5, 1988, pages 685-694, XP002045225 see page 686, right-hand column; figure 1 -----	1-11

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/06673

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0318341 A	31-05-1989	CA 1337048 A	19-09-1995
		AU 634987 B	11-03-1993
		AU 4495189 A	01-05-1990
		CA 2000661 A	14-04-1990
		WO 9004032 A	19-04-1990
		JP 3502644 T	20-06-1991
		US 5589615 A	31-12-1996
		AU 2811889 A	23-05-1989
		WO 8903887 A	05-05-1989
		EP 0319353 A	07-06-1989
		JP 2501802 T	21-06-1990
		US 5487991 A	30-01-1996
		AT 140959 T	15-08-1996
		DE 3855455 D	05-09-1996
		DE 3855455 T	09-01-1997
		AT 143694 T	15-10-1996
		DE 3855591 D	07-11-1996
		DE 3855591 T	03-04-1997
		EP 0723019 A	24-07-1996
		US 5623067 A	22-04-1997
WO 9410315 A	11-05-1994	AU 5446694 A	24-05-1994
WO 9416078 A	21-07-1994	AU 6162294 A	15-08-1994
		CA 2161881 A	21-07-1994
		EP 0745126 A	04-12-1996
WO 9515392 A	08-06-1995	AU 1255995 A	19-06-1995
		BR 9408228 A	26-08-1997
		CA 2177351 A	08-06-1995
		EP 0734445 A	02-10-1996
		HU 75078 A	28-04-1997
		JP 9511124 T	11-11-1997
		PL 314696 A	16-09-1996
		ZA 9409385 A	27-05-1996
EP 0319353 A	07-06-1989	CA 1337048 A	19-09-1995
		EP 0723019 A	24-07-1996
		AU 2811889 A	23-05-1989
		AU 634987 B	11-03-1993

# INTERNATIONAL SEARCH REPORT

information on patent family members

Int. J. Application No

PCT/US 98/06673

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0319353 A		AU 4495189 A	01-05-1990
		CA 2000661 A	14-04-1990
		WO 8903887 A	05-05-1989
		WO 9004032 A	19-04-1990
		EP 0318341 A	31-05-1989
		JP 2501802 T	21-06-1990
		JP 3502644 T	20-06-1991
		US 5487991 A	30-01-1996
		AT 143694 T	15-10-1996
		AT 140959 T	15-08-1996
		DE 3855455 D	05-09-1996
		DE 3855455 T	09-01-1997
		DE 3855591 D	07-11-1996
		DE 3855591 T	03-04-1997
		US 5623067 A	22-04-1997
-----			